



# Sur quelques algorithmes d'analyse de stabilité forte de matrices symplectiques

Mouhamadou Dosso

## ► To cite this version:

Mouhamadou Dosso. Sur quelques algorithmes d'analyse de stabilité forte de matrices symplectiques. Mathématiques [math]. Université de Bretagne occidentale - Brest, 2011. Français. NNT: 0CPZR600AC2 . tel-00626273

**HAL Id: tel-00626273**

**<https://theses.hal.science/tel-00626273>**

Submitted on 24 Sep 2011

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# THÈSE

présentée devant

**l'Université de Bretagne Occidentale**

pour obtenir le titre de :

Docteur de l'Université de Bretagne Occidentale  
Mention Mathématiques

par

**Mouhamadou Dosso**

École Doctorale SMIS  
Laboratoire de Mathématiques  
UFR Sciences et Techniques

Titre de la thèse :

**Sur quelques algorithmes d'analyse de stabilité  
forte de matrices symplectiques**

Soutenue le 25 septembre 2006 devant le jury composé de :

C. BREZINSKI	Professeur (Lille I)	Rapporteur
E. GALLOPOULOS	Professeur (Patras)	Rapporteur
R. BUCKDAHN	Professeur (Brest)	Examineur
A. SALAM	Maître de Conférences HDR (Calais)	Examineur
M. SADKANE	Professeur (Brest)	Dir. de thèse

# REMERCIEMENTS

Je tiens tout d'abord à exprimer toute ma gratitude à mon directeur de thèse, Monsieur Miloud Sadkane pour son aide précieuse, ses idées et ses conseils pendant mes quatre années de thèse.

Je remercie Messieurs Claude Brezinski et Efstratios Gallopoulos d'avoir accepté de rapporter sur ma thèse.

Je remercie également Messieurs Rainer Buckdahn et Ahmed Salam qui m'ont fait l'honneur de participer au Jury.

Mes remerciements vont également au ministère de l'enseignement supérieur et de la recherche scientifique de la Côte d'Ivoire d'avoir financé mes années de thèse.

J'adresse une pensée particulière à mes frères, soeurs et amis qui, malgré la distance, n'ont cessé de m'encourager.

Enfin, je remercie l'ensemble des enseignants chercheurs du département de Mathématiques de l'Université de Bretagne Occidentale ( Brest ) de m'avoir accueilli au sein de l'équipe.

# Table des matières

<b>1</b>	<b>Sur la stabilité des matrices symplectiques</b>	<b>3</b>
1.1	Introduction . . . . .	3
1.2	Quelques propriétés spectrales des matrices symplectiques . .	4
1.3	Stabilité des matrices symplectiques . . . . .	6
1.4	Stabilité des systèmes Hamiltoniens à coefficients périodiques	9
1.5	Une autre approche d'analyse de stabilité . . . . .	11
1.6	Travail effectué dans cette thèse . . . . .	13
<b>2</b>	<b>Dichotomie spectrale et matrices symplectiques</b>	<b>17</b>
2.1	Introduction . . . . .	17
2.2	Dichotomie spectrale d'une matrice par rapport à un cercle .	18
2.2.1	Une variante plus économique de l'algorithme 1 . . . .	23
2.2.2	Portrait spectral . . . . .	28
2.3	Application de l'algorithme aux matrices symplectiques . . .	29
2.4	Stabilité forte . . . . .	31
2.5	Conclusion . . . . .	42
<b>3</b>	<b>Une méthode de trichotomie spectrale de matrices symplectiques</b>	<b>45</b>
3.1	Introduction . . . . .	45
3.2	Notations et préliminaires techniques . . . . .	46
3.2.1	Principales notations . . . . .	46
3.2.2	Angles canoniques entre sous-espaces vectoriels . . . .	48
3.2.3	Définition de la convergence (sse) . . . . .	52
3.3	Premiers résultats de convergence . . . . .	60
3.3.1	Etude de la convergence de $(X_0^{(n)})_{n \geq 0}$ , $(X_1^{(n)})_{n \geq 0}$ et $(X_\infty^{(n)})_{n \geq 0}$	60
3.3.2	Convergence au sens (sse) . . . . .	65
3.4	Approximations de $\mathcal{X}_1$ et $\mathbb{P}_1$ . . . . .	68
3.4.1	Approximation de $\mathcal{X}_1$ . . . . .	69

3.4.2	Approximation de $\mathbb{P}_1$ . . . . .	75
3.5	Approximations des autres projecteurs . . . . .	78
3.6	Mise en œuvre informatique . . . . .	81
3.6.1	Calcul de $X_0^{(n)}$ , $X_1^{(n)}$ et $X_\infty^{(n)}$ . . . . .	81
3.6.2	Approximations des projecteurs . . . . .	88
3.6.3	Algorithme de trichotomie . . . . .	96
3.7	Une autre approche de détermination de la stabilité forte . . .	98
3.8	Illustrations numériques . . . . .	99
3.9	Conclusion . . . . .	102
	<b>Bibliographie</b>	<b>103</b>

# Table des figures

2.1	Portrait spectral (à gauche) et norme des résidus (à droite) sur la grille $l = 5$ .	29
2.2	Portrait spectral de $W_0$ (Exemple 2)	30
2.3	Partition du spectre de $W$ par des cercles d'équations $ \lambda - 1 / \lambda + 1  = cste$ .	33
2.4	Portrait spectral de $A$ (Exemple 3)	34
2.5	Distribution spectrale de $W(t)$ lorsque $t$ parcourt l'intervalle $[0, 2\pi]$	36
2.6	Portraits spectraux de $W$ (en haut à gauche) et de $A = (W - I)(W + I)^{-1}$ (en haut à droite) et la distribution spectrale de $W$ à $t = 0.3896$	37
2.7	Portraits spectraux de $W$ (en haut à gauche) et de $A = (W - I)(W + I)^{-1}$ (en haut à droite) et la distribution spectrale de $W$ à $t = 0.39$	39
2.8	Portraits spectraux de $W$ (en haut à gauche) et de $A = (W - I)(W + I)^{-1}$ (en haut à droite) et la distribution spectrale de $W$ à $t = 0.390017605$	41
2.9	Portraits spectraux de $W$ (en haut à gauche) et de $A = (W - I)(W + I)^{-1}$ (en haut à droite) et la distribution spectrale de $W$ à $t = 0.3905$	42



# Chapitre 1

## Sur la stabilité des matrices symplectiques

### 1.1 Introduction

Le but de ce chapitre est de rassembler quelques propriétés spectrales des matrices symplectiques dont on aura besoin tout le long de ce document. Cela permettra en particulier de fixer les notations et d'expliquer le sujet traité dans cette thèse. Un résumé du travail effectué dans cette thèse est donné à la fin de ce chapitre.

Soit  $J \in \mathbb{R}^{2N \times 2N}$  une matrice inversible et anti-symétrique.

**Définition 1** *On dit que  $W \in \mathbb{R}^{2N \times 2N}$  est  $J$ -symplectique si  $W^T J W = J$ . Parfois, on dira tout simplement que  $W$  est symplectique lorsque la matrice  $J$  est définie dans un contexte clair.*

Ces matrices font partie d'une classe importante de matrices dites "structurées" apparaissant généralement en théorie du contrôle [29, 16, 14, 3]. Les matrices symplectiques apparaissent notamment en contrôle optimal [14] et dans la théorie de résonance paramétrique [29]. En contrôle optimal, les matrices n'ont pas de valeurs propres sur le cercle unité, et, on est amené, dans ce domaine, à déterminer les projecteurs spectraux  $\mathbb{P}_0$  et  $\mathbb{P}_\infty$  sur les sous-espaces invariants (s-e-i) associés respectivement aux valeurs propres intérieures et extérieures au cercle unité. Dans le domaine de la résonance paramétrique, toutes les valeurs propres sont sur le cercle, et, on est amené à déterminer le projecteur spectral  $\mathbb{P}_1$  sur le s-e-i associé aux valeurs propres sur ce cercle. Dans ces domaines, les matrices symplectiques sont, en fait,



obtenues à partir de systèmes différentiels Hamiltoniens à coefficients périodiques (voir [29, Chap 3]), voir aussi la section 1.4

L'étude de la stabilité d'une matrice symplectique est étroitement liée à sa structure spectrale qui possède des propriétés intéressantes. Tout d'abord, le spectre d'une matrice symplectique réelle d'ordre  $2N$  est généralement composé de trois groupes :  $N_0$  valeurs propres intérieures au cercle unité,  $N_\infty = N_0$  valeurs propres extérieures au cercle unité et symétriquement placées par rapport au premier groupe, et  $N_1 = 2(N - N_0)$  valeurs propres sur le cercle unité. Les propriétés spectrales dont on aura besoin sont résumées ci-dessous.

## 1.2 Quelques propriétés spectrales des matrices symplectiques

Des propriétés simples et utiles pour le reste de cette thèse sont résumées dans le théorème suivant.

**Théorème 1** *Soit  $W \in \mathbb{R}^{2N \times 2N}$  une matrice  $J$ -symplectique. Alors*

1. *Les valeurs propres intérieures et extérieures au cercle unité sont symétriques au sens de l'inversion. Elles sont de la forme :*
  - .  $\lambda, \bar{\lambda}, 1/\lambda$  et  $1/\bar{\lambda}$  si  $\lambda \in \mathbb{C}$  avec  $|\lambda| \neq 1$  ;
  - .  $\lambda$  et  $\bar{\lambda}$  si  $\lambda \in \mathbb{C}$  avec  $|\lambda| = 1$  ;
  - .  $\lambda$  et  $1/\lambda$  si  $\lambda \in \mathbb{R}$ .
2. *Soient  $x$  et  $y$  deux vecteurs propres de  $W$  associés respectivement à des valeurs propres  $\lambda$  et  $\mu$ . Si  $\lambda\bar{\mu} \neq 1$ , alors  $y^* Jx := (Jx, y) = 0$ . En particulier, si  $|\lambda| \neq 1$ ,  $(Jx, x) = 0$  pour tout  $x \in \text{Ker}(W - \lambda I)$ . Et plus généralement on a :*
3. *Soient  $X \in \mathbb{C}^{2N \times k_1}$  et  $Y \in \mathbb{C}^{2N \times k_2}$  deux matrices dont les colonnes engendrent des s-e-i de  $W$  :*

$$WX = XA, \quad WY = YB. \quad (1.1)$$

*Si les valeurs propres  $\lambda_i(A)$  et  $\lambda_j(B)$  de  $A$  et  $B$  sont telles que  $\lambda_i(A)\bar{\lambda}_j(B) \neq 1$  pour tout  $i$  et  $j$ , alors  $X^* JY = 0$ .*

4. *Si  $\lambda$  est une valeur propre défective (i.e. non semi-simple) sur le cercle unité, alors il existe un vecteur propre  $x$  associé à  $\lambda$  tel que  $(Jx, x) = 0$ .*
5. *Il existe  $T_1 \in \mathbb{R}^{2N \times 2N}$  inversible et trois matrices  $W_\infty \in \mathbb{R}^{N_\infty \times N_\infty}$ ,  $W_1 \in \mathbb{R}^{N_1 \times N_1}$  et  $W_0 \in \mathbb{R}^{N_0 \times N_0}$  avec  $N_\infty = N_0 \leq N$  et  $N_1 = 2(N - N_0)$  dont les*

valeurs propres sont respectivement extérieures au cercle unité, sur le cercle unité et intérieures au cercle unité, et telles que :

$$T_1^{-1}WT_1 = \begin{pmatrix} W_\infty & 0 & 0 \\ 0 & W_1 & 0 \\ 0 & 0 & W_0 \end{pmatrix}.$$

De plus

$$T_1^T J T_1 = \begin{pmatrix} 0 & 0 & -M^T \\ 0 & J_1 & 0 \\ M & 0 & 0 \end{pmatrix},$$

où  $M$  est inversible et  $J_1$  est anti-symétrique et inversible.

6. Dans le cas particulier (mais fréquemment utilisé) où  $J = \begin{pmatrix} 0 & -I_N \\ I_N & 0 \end{pmatrix}$  ( $I_N$  désigne la matrice d'identité d'ordre  $N$ ), la matrice  $W$  a une décomposition en valeur singulière de la forme

$$W = U \begin{pmatrix} \Omega & 0 \\ 0 & \Omega^{-1} \end{pmatrix} V^T, \quad (1.2)$$

où  $U$  et  $V$  sont orthogonales et  $J$ -symplectiques et  $\Omega = \text{diag}(\omega_1, \dots, \omega_N)$  avec  $\omega_1 \geq \dots \geq \omega_N \geq 1$ .

### Démonstration

Les propriétés 1 - 5 se trouvent par exemple dans [10] et la propriété 6 dans [28].  $\square$

Exemples de matrices  $J$ -symplectiques : Les trois premiers exemples sont valables pour  $J = \begin{pmatrix} 0 & -I_N \\ I_N & 0 \end{pmatrix} \in \mathbb{R}^{2N \times 2N}$  :

1. Si  $X \in \mathbb{R}^{N \times N}$  est inversible, alors  $\begin{pmatrix} X & 0 \\ 0 & X^{-T} \end{pmatrix}$  est  $J$ -symplectique.
2. Si  $Y \in \mathbb{R}^{N \times N}$  est symétrique, alors la matrice  $\begin{pmatrix} I_N & Y \\ 0 & I_N \end{pmatrix}$  est  $J$ -symplectique.
3. Si  $X, Y \in \mathbb{R}^{N \times N}$  satisfont  $XY^T = YX^T$  et  $X$  est inversible, alors la matrice  $\begin{pmatrix} X & Y \\ 0 & X^{-T} \end{pmatrix}$  est  $J$ -symplectique.
4. Si  $W \in \mathbb{R}^{2N \times 2N}$  est  $J$ -symplectique, alors  $W^T$ ,  $W^{-T}$ ,  $W^{-1}$  sont  $J$ -symplectiques.
5. Le produit de matrices  $J$ -symplectiques est  $J$ -symplectique.

### 1.3 Stabilité des matrices symplectiques

Soit  $W$  une matrice  $J$ -symplectique. Les définitions classiques de stabilité sont données ci-dessous. Sauf mention contraire, le symbole  $\| \cdot \|$  désigne la norme spectrale (norme 2).

**Définition 2** *On dit que  $W$  est stable si  $\|W^k\| < \infty$  pour tout  $k \in \mathbb{N}$  et pour toute norme matricielle  $\| \cdot \|$ .*

Puisque  $W^T J W = J$ , la condition  $\|W^k\| < \infty$  est équivalente à  $\|W^{-k}\| < \infty$ .

Il est clair qu'une matrice symplectique est stable si et seulement si toutes ses valeurs propres sont sur le cercle unité et sont non défectives. La définition suivante caractérise un autre type de stabilité : la stabilité forte.

**Définition 3** *On dit que  $W$  est fortement stable s'il existe  $\epsilon > 0$  tel que toute matrice  $\widetilde{W}$   $J$ -symplectique vérifiant  $\|W - \widetilde{W}\| \leq \epsilon$  est stable.*

En d'autres termes,  $W$  est fortement stable si elle est stable et le reste quand elle est soumise à des petites perturbations conservant la symplecticité. Ce concept est plus robuste que celui de la stabilité. Il est aussi plus délicat à caractériser. Bien entendu, les valeurs propres d'une telle matrice doivent être sur le cercle unité et semi-simples, mais cela ne suffit pas à caractériser ce type de stabilité. En effet, les valeurs propres doivent, en plus, être soit de type 1 ou de type 2 conformément à la définition suivante :

**Définition 4** *Soit  $\lambda$  une valeur propre de module 1 de  $W$ . On dit que  $\lambda$  est de type 1 (resp. de type 2) si tout vecteur propre  $x$  associé à  $\lambda$  vérifie  $(iJx, x) > 0$  (resp.  $(iJx, x) < 0$ ). Si  $(Jx, x) = 0$ , on dira que  $\lambda$  est de type mêlé.*

**Remarque :** La définition ci-dessus a bien un sens puisque la matrice  $iJ$  est complexe Hermitienne.

Krein, Gelfand et Lidskii [29] ont introduit cette classification et l'ont utilisée pour caractériser de manière élégante la stabilité forte de matrices symplectiques. La caractérisation est résumée dans le théorème suivant, que nous appellerons par la suite : le critère de Krein, Gelfand et Lidskii (critère KGL).

**Théorème 2 (critère KGL)**

*La matrice  $W$  est fortement stable si et seulement si toutes ses valeurs propres sont sur le cercle unité et sont soit de type 1 ou de type 2.*

Ce théorème est obtenu comme le résultat de plusieurs autres théorèmes "éparpillés" dans [29]. Nous proposons un résumé des étapes les plus importantes pour sa démonstration.

Remarquons d'abord que d'après le point 2 du théorème 1, tout vecteur propre  $x$  de  $W$  associé à une valeur propre  $\lambda$  en dehors du cercle unité satisfait  $(Jx, x) = 0$ . Ceci permet d'étendre la définition d'une valeur propre de type mêlé aux valeurs propres en dehors du cercle unité.

Lorsque  $W$  n'a pas de valeurs propres de type mêlé, on a le théorème suivant.

**Théorème 3** *Soit  $\lambda$  une valeur propre de type non mêlé de  $W$ , i.e.  $\lambda$  est sur le cercle unité et est soit de type 1 ou de type 2. Alors, il existe des constantes  $\delta > 0$  et  $\gamma > 0$  telles que pour toute matrice  $J$ -symplectique  $\widetilde{W}$  satisfaisant  $\|W - \widetilde{W}\| < \delta$ , les valeurs propres  $\tilde{\lambda}$  de  $\widetilde{W}$  dans le voisinage  $|\lambda - \tilde{\lambda}| < \gamma$  sont sur le cercle unité et sont semi-simples.*

### Démonstration

Elle se trouve dans [29, p. 161]. Nous la reproduisons pour plus de clarté. Raisonnons par l'absurde : il existe alors une suite de matrices  $W_1, W_2, \dots \rightarrow W$  de valeurs propres respectives  $\lambda_1, \lambda_2, \dots, \rightarrow \lambda$  telles que  $\lambda_m$  soit une valeur propre de  $W_m$  avec  $|\lambda_m| \neq 1$  où  $\lambda$  est de module 1 mais défective (i.e. non semi-simple). Soit  $x_m$  un vecteur propre correspondant à  $\lambda_m$  tel que  $\|x_m\| = 1$ . Le théorème 1 donne  $(Jx_m, x_m) = 0$ ,  $m = 1, 2, 3, \dots$ . La suite  $(x_m)$  étant bornée, on peut en extraire une sous-suite convergente vers un vecteur  $x$ , et par passage à la limite, on obtient  $Wx = \lambda x$ , et  $\|x\| = 1$ ,  $(Jx, x) = 0$ . La valeur propre  $\lambda$  est donc de type mêlé. D'où une contradiction.  $\square$

Pour les valeurs propres de type mêlé, nous proposons d'abord le résultat suivant :

**Lemme 1** *Les matrices*

$$\widetilde{W}_0 = K^{-1} \begin{pmatrix} W_0 & 0 \\ 0 & W_0^T \end{pmatrix} K \quad \text{et} \quad \widetilde{J}_0 = K^T \begin{pmatrix} J_0 & 0 \\ 0 & J_0^T \end{pmatrix} K,$$

où  $K$  est inversible,  $W_0 = \begin{pmatrix} \cos \varphi & -\sin \varphi \\ \sin \varphi & \cos \varphi \end{pmatrix}$ ,  $J_0 = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$  vérifient les propriétés suivantes :

1.  $\widetilde{W}_0$  est  $\widetilde{J}_0$ -symplectique. Ses valeurs propres, égales à  $e^{\pm i\varphi}$ , sont de type mêlé.

2. Il existe une matrice  $\widetilde{W}(\alpha)$   $\widetilde{J}_0$ -symplectique, de valeurs propres  $e^{\pm\alpha\pm i\varphi}$ , et telle que  $\lim_{\alpha\rightarrow 0} \|\widetilde{W}_\alpha - \widetilde{W}_0\| = 0$ .

### Démonstration

Il est facile de vérifier le point 1. Pour le point 2, posons  $I = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$  et

$I_0 = \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix}$  et remarquons que :

$$\begin{aligned}\widetilde{J}_0 &= K^T \begin{pmatrix} I & 0 \\ 0 & I_0 \end{pmatrix} \begin{pmatrix} J_0 & 0 \\ 0 & -J_0 \end{pmatrix} \begin{pmatrix} I & 0 \\ 0 & I_0 \end{pmatrix} K \\ \widetilde{W}_0 &= K^{-1} \begin{pmatrix} I & 0 \\ 0 & I_0 \end{pmatrix} \begin{pmatrix} W_0 & 0 \\ 0 & W_0 \end{pmatrix} \begin{pmatrix} I & 0 \\ 0 & I_0 \end{pmatrix} K.\end{aligned}$$

Posons

$$\widetilde{W}_1(\alpha) = K^{-1} \begin{pmatrix} I & 0 \\ 0 & I_0 \end{pmatrix} \begin{pmatrix} \operatorname{ch}(\alpha)I & \operatorname{sh}(\alpha)I \\ \operatorname{sh}(\alpha)I & \operatorname{ch}(\alpha)I \end{pmatrix} \begin{pmatrix} I & 0 \\ 0 & I_0 \end{pmatrix} K.$$

On vérifie facilement que  $\widetilde{W}_1(\alpha)$  est  $\widetilde{J}_0$ -symplectique et par conséquent le produit  $\widetilde{W}(\alpha) = \widetilde{W}_0 \widetilde{W}_1(\alpha)$  l'est aussi. Et on a

$$\widetilde{W}(\alpha) = K^{-1} \begin{pmatrix} I & 0 \\ 0 & I_0 \end{pmatrix} \begin{pmatrix} W_0 & 0 \\ 0 & W_0 \end{pmatrix} \begin{pmatrix} \operatorname{ch}(\alpha)I & \operatorname{sh}(\alpha)I \\ \operatorname{sh}(\alpha)I & \operatorname{ch}(\alpha)I \end{pmatrix} \begin{pmatrix} I & 0 \\ 0 & I_0 \end{pmatrix} K.$$

Les valeurs propres de  $\widetilde{W}(\alpha)$  sont donc les mêmes que celles de

$$\begin{pmatrix} W_0 & 0 \\ 0 & W_0 \end{pmatrix} \begin{pmatrix} \operatorname{ch}(\alpha)I & \operatorname{sh}(\alpha)I \\ \operatorname{sh}(\alpha)I & \operatorname{ch}(\alpha)I \end{pmatrix}$$

qui sont égales à  $e^{\pm\alpha\pm i\varphi}$ . De plus,  $\widetilde{W}(\alpha) - \widetilde{W}_0$  se décompose sous la forme

$$K^{-1} \begin{pmatrix} I & 0 \\ 0 & I_0 \end{pmatrix} \begin{pmatrix} W_0 & 0 \\ 0 & W_0 \end{pmatrix} \begin{pmatrix} (\operatorname{ch}(\alpha) - 1)I & \operatorname{sh}(\alpha)I \\ \operatorname{sh}(\alpha)I & (\operatorname{ch}(\alpha) - 1)I \end{pmatrix} \begin{pmatrix} I & 0 \\ 0 & I_0 \end{pmatrix} K.$$

Il est clair que  $\lim_{\alpha\rightarrow 0} \|\widetilde{W}_\alpha - \widetilde{W}_0\| = 0$ .  $\square$

Ce lemme se généralise de la manière suivante (voir [29, p. 192-195]).

**Théorème 4** *Si  $\lambda$  est une valeur propre de type mêlé de la matrice  $J$ -symplectique  $W$ , alors il existe une matrice  $\widetilde{W}$   $J$ -symplectique arbitrairement proche de  $W$  ayant des valeurs propres extérieures et intérieures au cercle unité dans un voisinage de  $\lambda$ .*

On peut à présent montrer le théorème 2 (critère KGL) :

Si  $W$  est fortement stable, il est clair que toutes ses valeurs propres sont sur le cercle unité. Ces valeurs propres ne peuvent être de type mêlé car sinon, d'après le théorème 4,  $W$  serait arbitrairement proche d'une matrice  $\widetilde{W}$   $J$ -symplectique ayant des valeurs extérieures et intérieures au cercle unité, ce qui contredit la stabilité de  $W$ .

Réciproquement, si toutes les valeurs propres de  $W$  sont sur le cercle unité et sont soit de type 1 ou de type 2, le théorème 3 montre que les valeurs propres des petites perturbations  $\widetilde{W}$   $J$ -symplectiques de  $W$  sont sur le cercle unité et semi-simples. Autrement dit,  $W$  est fortement stable.  $\square$

Comme nous l'avons annoncé dans l'introduction, les matrices symplectiques sont souvent obtenues comme solutions de systèmes Hamiltoniens à coefficients périodiques. Nous allons à présent préciser ce point.

## 1.4 Stabilité des systèmes Hamiltoniens à coefficients périodiques

Un système linéaire Hamiltonien à coefficients  $T$ -périodiques est un système différentiel de la forme

$$J \frac{dx(t)}{dt} = H(t)x(t), \quad t \in \mathbb{R} \quad (1.3)$$

où  $H(t) \in \mathbb{R}^{2N \times 2N}$  est symétrique et  $T$ -périodique, i.e.  $H(t+T) = H(t) = (H(t))^T$ . On sait (voir par exemple [29, Chap 3.]) que la solution fondamentale  $X(t)$  de (1.3) (i.e. la matrice  $X(t)$  satisfaisant  $J \frac{dX(t)}{dt} = H(t)X(t)$  et  $X(0) = I$ ) est  $J$ -symplectique et satisfait pour tout  $t \in \mathbb{R}$  la relation

$$X(t+T) = X(t)X(T) \quad (\neq X(T)X(t)). \quad (1.4)$$

Concernant la stabilité de tels systèmes, on a d'abord la définition classique suivante :

**Définition 5** *Le système (1.3) est stable si chacune de ses solutions  $x(t)$  reste bornée pour  $t \in \mathbb{R}$ .*

En vertu de (1.4) on a pour tout  $t$  et tout  $n \in \mathbb{N}$

$$X(t+nT) = X(t)X^n(T). \quad (1.5)$$

Par conséquent toute solution  $x(t) = X(t)x_0$  de (1.3) vérifie

$$\sup_{t \in \mathbb{R}} \|x(t)\| \leq \sup_{0 \leq t \leq T} \|X(t)\| \sup_{n \in \mathbb{N}} \|X^n(T)x_0\|.$$

Et d'autre part

$$\begin{aligned} \sup_{n \in \mathbb{N}} \|X^n(T)x_0\| &= \sup_{n \in \mathbb{N}} \|X(nT)x_0\| \\ &= \sup_{n \in \mathbb{N}} \|x(nT)\| \leq \sup_{t \in \mathbb{R}} \|x(t)\|. \end{aligned}$$

Autrement dit, une condition nécessaire et suffisante de stabilité de (1.3) est que les puissances de  $X(T)$  restent bornées. Et puisque  $X(T)$  est  $J$ -symplectique cela signifie (cf. définition 2 du Chapitre 1) qu'il existe une constante  $C$  telle que

$$\sup_{n \in \mathbb{Z}} \|X^n(T)\| \leq C.$$

La stabilité de (1.3) se ramène donc à celle de la matrice  $J$ -symplectique  $X(T)$ . On sait, d'après le chapitre 1, que cela équivaut au fait que toutes les valeurs propres de  $X(T)$  soient sur le cercle unité et semi-simples. Mais ces conditions ne suffisent pas à garantir que les solutions restent bornées lorsque le système (1.3) est soumis à des perturbations Hamiltoniennes. C'est l'objet de la stabilité forte dont on rappelle la définition (voir [29, Chap.3], [12]).

**Définition 6** *Le système (1.3) est fortement stable si tout système Hamiltonien à coefficients  $T$ -périodiques assez voisin de (1.3) est stable.*

Plus précisément, le système (1.3) est fortement stable s'il existe  $\epsilon > 0$  tel que tout système Hamiltonien à coefficients  $T$ -périodiques de la forme

$$J \frac{dx(t)}{dt} = \tilde{H}(t)x(t)$$

et satisfaisant

$$\|H - \tilde{H}\| \equiv \int_0^T \|H(t) - \tilde{H}(t)\| dt < \epsilon$$

est stable. Et on a le théorème important suivant (voir [29, p. 196]) :

**Théorème 5** *Le système (1.3) est fortement stable si et seulement si la matrice  $J$ -symplectique  $X(T)$  est fortement stable.*

Ainsi, la stabilité (resp. la forte stabilité) de (1.3) se ramène à la stabilité (resp. la forte stabilité) de la matrice  $J$ -symplectique  $W = X(T)$  qui est la solution fondamentale de (1.3) évaluée en la période  $T$ . C'est la raison pour laquelle nous concentrons nos efforts dans cette thèse sur la construction d'algorithmes permettant d'analyser la stabilité forte de matrices symplectiques.

## 1.5 Une autre approche d'analyse de stabilité

L'analyse de la stabilité (stabilité forte) de ces matrices symplectiques passe donc par le calcul des valeurs propres de matrices non symétriques et la vérification du critère KGL. Ces valeurs propres sont a priori inconnues et leur calcul numérique peut être sensible aux inévitables erreurs d'arrondis des méthodes utilisées. C'est le pseudo-spectre que l'on calcule par des méthodes numériques [27, 8], i.e. le spectre de perturbations de  $W$ . Il est bien connu qu'à l'exception du cas symétrique ou normal, ce qui n'est pas notre cas, le pseudo-spectre d'une matrice peut être différent de son spectre (voir [27, 8]). Par conséquent, le critère KGL, excellent outil mathématique, peut, numériquement, induire en erreur. Du point de vue numérique, il est préférable de changer la classification donnée dans la définition 4. Pour cela on aura besoin d'une autre classification introduite à partir de la matrice suivante (voir [5, 10])

$$S_0 = \frac{1}{2}J(W - W^{-1}). \quad (1.6)$$

La matrice  $S_0$  possède quelques propriétés simples et utiles. Par exemple :

### Proposition 1

1. La matrice  $S_0$  est symétrique et satisfait  $W^T S_0 W = S_0$ .
2. Si  $S_0$  est définie positive ( $S_0 \geq 0$ ) ou définie négative ( $S_0 \leq 0$ ), alors les valeurs propres de  $W$  sont sur le cercle unité.

### Démonstration

La matrice  $S_0$  est symétrique car  $W$  est  $J$ -symplectique :  $S_0 = \frac{1}{2}(JW + (JW)^T) = S_0^T$ . La propriété  $W^T S_0 W = S_0$  se vérifie facilement et entraîne le point 2 car  $(S_0 x, x) = (W^T S_0 W x, x) = (S_0 W x, W x)$ .  $\square$

Remarquons que  $S_0$  est singulière si et seulement si  $W$  possède les valeurs propres 1 ou  $-1$ . De plus, si  $(\lambda, x)$  est un élément propre de  $W$  avec  $\lambda = e^{i\theta}$ ,  $0 < \theta < \pi$ , alors

$$(S_0 x, x) = \sin \theta (i J x, x). \quad (1.7)$$

Par conséquent, si  $(S_0 x, x) > 0$  alors  $\lambda = e^{i|\theta|}$  est une valeur propre de type 1 et  $\lambda = e^{-i|\theta|}$  est de type 2, et si  $(S_0 x, x) < 0$  alors  $\lambda = e^{-i|\theta|}$  est de type 1 et  $\lambda = e^{i|\theta|}$  est de type 2. Si  $\theta = 0$  ou  $\theta = \pi$  alors  $\lambda = \pm 1$ . De telles valeurs propres sont nécessairement de type mêlé car le vecteur propre correspondant  $x$  est réel et satisfait  $(Jx, x) = (x, J^T x) = -(x, Jx) = -(Jx, x)$ . Donc  $(Jx, x) = 0$  et  $(S_0 x, x) = 0$ . Par conséquent, si le critère KGL est satisfait (il n'y a pas de valeurs propres de type mêlé), alors  $S_0$  est inversible. Ceci



conduit à une autre classification des valeurs propres. Au lieu de les classer en type 1 et 2, nous leur attribuerons deux couleurs : rouge ( $r$ -valeur propre) et verte ( $v$ -valeur propre) conformément à la définition suivante :

**Définition 7** *Soit  $\lambda$  une valeur propre de module 1 de  $W$ . On dit que  $\lambda$  est une  $r$ -valeur propre (resp.  $v$ -valeur propre) si tout vecteur propre  $x$  associé à  $\lambda$  vérifie  $(S_0x, x) > 0$  (resp.  $(S_0x, x) < 0$ ).*

Comme on la verra par la suite, cette classification est plus appropriée au calcul numérique que celle donnée dans la définition 4. En effet, elle utilise uniquement des matrices symétriques. La différence principale entre les définitions 4 et 7 est la suivante : si  $\lambda = e^{i|\theta|}$  et  $\bar{\lambda} = e^{-i|\theta|}$  sont des valeurs propres de  $W$  respectivement de type 1 et 2, alors tout vecteur  $z$  de la forme  $z = \alpha x + \beta \bar{x}$  avec  $Wx = \lambda x$  et  $W\bar{x} = \bar{\lambda}\bar{x}$  vérifie

$$(S_0z, z) = 2 \sin |\theta| (|\alpha|^2 (iJx, x) - |\beta|^2 (iJ\bar{x}, \bar{x})),$$

et puisque  $(iJx, x) > 0$  et  $(iJ\bar{x}, \bar{x}) < 0$ , on en déduit que la forme quadratique  $(S_0z, z)$  conserve le même signe pour tout  $z$  dans le s-e-i associé aux valeurs propres  $\lambda$  et  $\bar{\lambda}$ . Par conséquent, les valeurs propres  $\lambda$  et  $\bar{\lambda}$  auront la même couleur (rouge ou verte). D'autre part, un vecteur propre  $x$  est associé à une valeur propre de type mêlé si et seulement si  $(Jx, x) = 0$ , qui, d'après (1.7), s'écrit aussi  $(S_0x, x) = 0$  (remarquons que le cas où  $\sin \theta = 0$  conduit aux valeurs propres de type mêlé  $\pm 1$ , et donc  $(Jx, x) = 0 = (S_0x, x)$ ). Donc, la définition 7 ne change pas les valeurs propres de type mêlé.

En se basant sur la discussion ci-dessus, on obtient les résultats suivants.

**Proposition 2** *Le spectre d'une matrice symplectique  $W$  est formé uniquement de  $r$ - et/ou de  $v$ -valeurs propres si et seulement si il est formé uniquement de valeurs propres de type 1 ou 2.*

**Théorème 6** *La matrice  $W$  est fortement stable si et seulement si :*

1. toutes ses valeurs propres sont sur le cercle unité ;
2. les valeurs propres sont soit de couleur rouge ( $r$ -valeurs propres) soit de couleur verte ( $v$ -valeurs propres).

Du point de vue pratique, il faut en plus des deux conditions ci-dessus, imposer que les  $r$ - et  $v$ -valeurs propres soient bien séparées.

Si on désigne par  $\mathbb{P}_r$  le projecteur sur le sous-espace invariant associé à toutes les  $r$ -valeurs propres de  $W$  et par  $\mathbb{P}_v$  celui associé aux  $v$ -valeurs propres, alors on a le théorème suivant qui est une re-écriture du théorème précédent.

**Théorème 7** *La matrice  $W$  est fortement stable si et seulement si*

$$\mathbb{P}_r + \mathbb{P}_v = \mathbb{P}_1 = I_{2N}. \quad (1.8)$$

Tous les algorithmes développés dans cette thèse pour analyser la stabilité forte sont basés sur le théorème 7.

## 1.6 Travail effectué dans cette thèse

Nous résumons brièvement le travail effectué dans cette thèse.

Dans le chapitre 2, nous adaptons les méthodes de dichotomie spectrale à des matrices symplectiques. De façon générale, ces méthodes permettent de décider s'il y a ou non des valeurs propres d'une matrice sur ou dans un voisinage d'un contour  $\gamma$  donné du plan complexe. Dans le cas où de telles valeurs propres n'existent pas, ces méthodes calculent le projecteur spectral sur le sous-espace invariant (s-e-i) associé aux valeurs propres intérieures à  $\gamma$ . Le calcul de ce projecteur est accompagné par une quantité appelée critère de dichotomie. Cette quantité représente un indicateur de confiance associé au calcul du projecteur. Plus elle est petite, meilleure est la qualité numérique du projecteur. Tout d'abord, nous proposons un nouvel algorithme de dichotomie spectrale d'une matrice par un cercle (Algorithme 2 du Chapitre 2). Cet algorithme est une variante économique d'un algorithme proposé dans [11]. Il possède les mêmes propriétés mathématiques que ce dernier mais requiert quatre fois moins d'opérations. Ensuite, nous adaptons cet algorithme au cas des matrices symplectiques en utilisant l'approche suivante : en examinant le portrait spectral de la matrice symplectique  $W$ , i.e. le graphe de la fonction  $r \mapsto \|H(r)\|$ , où

$$H(r) = \frac{1}{2\pi} \int_0^{2\pi} \left( I - e^{i\theta} \frac{W^*}{r} \right)^{-1} \left( I - e^{-i\theta} \frac{W}{r} \right)^{-1} d\theta,$$

on en déduit un paramètre  $r_0$  tel que le cercle  $C(0, r_0)$  de centre 0 et de rayon  $r_0$  contienne toutes les valeurs propres de  $W$  dont le module est strictement inférieur à 1 et excluant les autres. L'algorithme proposé (Algorithme 2) permet alors de calculer le projecteur spectral

$$\mathbb{P}_0 = \frac{1}{2i\pi} \int_{C(0, r_0)} (zI - W)^{-1} dz$$

sur le s-e-i associé aux valeurs propres dans  $C(0, r_0)$ . Puis, en utilisant la symplecticité de  $W$  (symétrie du spectre par rapport au cercle unité), on en

déduit un autre paramètre  $r_\infty = 1/r_0$  et le projecteur spectral  $\mathbb{P}_\infty$  sur le s-e-i associé aux valeurs propres de module supérieur à 1. Et, enfin, le projecteur spectral  $\mathbb{P}_1 = I - \mathbb{P}_0 - \mathbb{P}_\infty$  sur le s-e-i associé aux valeurs propres sur le cercle unité.

Lorsque  $\mathbb{P}_0 = \mathbb{P}_\infty = 0$  (et donc  $\mathbb{P}_1 = I$ ) et  $W$  n'a pas les valeurs propres (mêlées)  $-1$  et  $1$ , nous proposons une étude basée encore sur l'algorithme 2, et inspirée de [5], pour analyser la stabilité forte. L'idée est de partitionner, s'il y a stabilité forte, le spectre de  $W$  en deux parties disjointes constituées uniquement des  $r$ - et/ou  $v$ -valeurs propres. Pour cela, on se sert de la transformation  $\lambda \mapsto (\lambda - 1)/(\lambda + 1)$  qui permet de "recouvrir" le spectre de  $W$  en plusieurs cercles (comme en figure 2.3 du chapitre 2). Plus précisément, on considère la matrice  $A = (W + I)^{-1}(W - I)$  qui a les mêmes s-e-i que  $W$ . Ses valeurs propres sont de la forme  $l_k = (\lambda_k - 1)/(\lambda_k + 1)$ , où  $\lambda_k$  désigne une valeur propre de  $W$  de module 1. En examinant le portrait spectral de  $A$ , on en déduit des nombres positifs  $a_1, a_2, \dots$  qui entrelacent strictement les modules  $|l_1|, |l_2|, \dots$ . L'algorithme 2 permet de calculer les projecteurs  $P_1, P_2, \dots$  sur les s-e-i associés aux valeurs propres de  $A$  dans les cercles  $C(0, a_k)$ , et par suite les projecteurs  $Q_k = P_{k+1} - P_k$  sur les s-e-i de  $A$  (et donc de  $W$ ) associés aux valeurs propres  $l = (\lambda - 1)/(\lambda + 1)$  de  $A$  situées dans des couronnes de la forme  $\{\lambda \in \mathbb{C} : a_k < |\lambda - 1|/|\lambda + 1| < a_{k+1}\}$ . Si on note  $S_k = Q_k^T S_0 Q_k \equiv S_k^T$ , alors les projecteurs (supplémentaires)  $\mathbb{P}_r$  et  $\mathbb{P}_v$  sur les s-e-i associés aux  $r$ - et  $v$ -valeurs propres sont donnés par  $\mathbb{P}_r + \mathbb{P}_v = \mathbb{P}_1 \equiv I$  avec

$$\mathbb{P}_r = \sum_{S_k \geq 0} Q_k \quad \text{et} \quad \mathbb{P}_v = \sum_{S_k \leq 0} Q_k.$$

L'algorithme 3 du chapitre 2 résume les différentes étapes permettant le calcul de ces projecteurs.

Dans le chapitre 3 nous proposons et analysons la convergence d'un algorithme de trichotomie spectrale de matrices symplectiques (Algorithme 4). Cet algorithme utilise une approche différente de celle du chapitre 2. Il permet de s'affranchir du choix du paramètre  $r_0$  qui peut-être coûteux et difficile à déterminer.

En notant par  $\mathcal{X}_0, \mathcal{X}_1, \mathcal{X}_\infty$ , les sous-espaces invariants associés aux valeurs propres de  $W$  intérieures, sur, et en dehors du cercle unité, l'algorithme proposé permet de construire des suites de sous-espaces  $(\mathcal{X}_0^{(n)})_{n \geq 0}, (\mathcal{X}_1^{(n)})_{n \geq 0}$  et  $(\mathcal{X}_\infty^{(n)})_{n \geq 0}$  qui convergent, en un sens que nous précisons, respectivement vers  $\mathcal{X}_0 + \mathcal{X}_1^-, \mathcal{X}_1$  et  $\mathcal{X}_\infty + \mathcal{X}_1$ . Ensuite, une étape dite de J-orthogonalisation permet d'éliminer les parties correspondantes au sous-espace  $\mathcal{X}_1$  dans  $\mathcal{X}_0^{(n)}$  et

$\mathcal{X}_\infty^{(n)}$ . L'algorithme permet finalement de bloc-diagonaliser  $W$  sous la forme :

$$W = X \begin{pmatrix} W_0 & & \\ & W_1 & \\ & & W_\infty \end{pmatrix} X^{-1},$$

où  $W_0$ ,  $W_1$  et  $W_\infty$  sont des matrices dont les valeurs propres sont respectivement intérieures, sur, et en dehors du cercle unité, ainsi qu'une approximation des projecteurs spectraux  $\mathbb{P}_0$ ,  $\mathbb{P}_1$  et  $\mathbb{P}_\infty$ .

Lorsque  $\mathbb{P}_0 = \mathbb{P}_\infty = 0$ , nous utilisons une approche inspirée de [6] pour analyser la stabilité forte. L'idée est de construire de manière itérative la suite

$$S^{(n)} = \frac{1}{2^n} \sum_{k=1}^{2^n} (W^T)^{k-1} W^{k-1},$$

qui, lorsque  $W$  est stable, converge vers une matrice symétrique définie positive  $S$  satisfaisant  $W^T S W = S$ . En pratique, pour analyser la stabilité forte, on considère, pour  $n$  assez grand, les valeurs propres du faisceau symétrique  $S_0 - \lambda S^{(n)}$ . Lorsqu'il y a stabilité forte, les valeurs propres de ce faisceau sont réelles et non nulles. Les projecteurs  $\mathbb{P}_r$  et  $\mathbb{P}_v$  sont obtenus à partir des s-e-i de ce faisceau associés respectivement aux valeurs strictement positives et strictement négatives.



## Chapitre 2

# Dichotomie spectrale et matrices symplectiques

### 2.1 Introduction

Dans ce chapitre nous adaptons les méthodes de dichotomie spectrale à des matrices symplectiques. Le but final est de construire un algorithme de type "dichotomie spectrale" permettant d'analyser la stabilité forte d'une matrice  $W \in \mathbb{R}^{2N \times 2N}$ ,  $J$ -symplectique. Rappelons qu'une telle matrice est fortement stable si et seulement si elle satisfait le critère KGL (cf. théorème 2 du chapitre 1) ou, de manière équivalente, son spectre est sur le cercle unité et est constitué uniquement de valeurs propres de couleur rouge et/ou verte (cf. théorème 5 du chapitre 1). Notre analyse sera basée sur ce dernier théorème qui, comme nous l'avons déjà mentionné, est plus propice au calcul numérique.

Les méthodes de dichotomie spectrale ont été introduites par Godunov [4]. Elles permettent de décider s'il y a ou non des valeurs propres sur ou dans un voisinage d'un contour  $\gamma$  du plan complexe. Dans le cas où de telles valeurs propres n'existent pas, ces méthodes calculent le projecteur spectral  $\mathbb{P}$  sur le s-e-i associé aux valeurs propres intérieures (extérieures) à  $\gamma$ . Le calcul du projecteur est accompagné par celui de la norme  $\|H\|_2$  d'une matrice Hermitienne définie positive, appelée critère de dichotomie. Cette quantité représente un indicateur de confiance associé au calcul du projecteur. Plus elle est petite, meilleure est la qualité numérique du projecteur. Outre leur robustesse, ces méthodes ont l'avantage de se prêter facilement au calcul sur machines parallèles [1]. Le cas où  $\gamma$  est un cercle est traité dans [4]. Des

généralisations pour des faisceaux matriciels et aux cas où  $\gamma$  est un cercle, une ellipse ou un polygone ont été développées par Godunov, Malyshev et Sadkane dans [18, 19, 7, 20, 9].

Dans ce chapitre, nous nous intéressons à la dichotomie spectrale d'une matrice par rapport à un cercle. Ceci est naturel puisque notre objectif est d'adapter la méthode à des matrices symplectiques et de faire des caractéristiques spectrales de ces dernières (cf. Chapitre 1). Un algorithme de dichotomie spectrale a été récemment proposé par Godunov et Sadkane dans [11]. Cet algorithme calcule le projecteur  $\mathbb{P}$  et la matrice  $H$  de manière itérative<sup>1</sup>. Il requiert, en revanche, un nombre d'opérations arithmétiques relativement élevé. De ce fait, nous développons une variante de cet algorithme qui possède les mêmes propriétés mathématiques que l'algorithme proposé dans [11] mais dont le coût en nombre d'opérations est divisé par quatre.

La section 2 est consacrée à l'élaboration de ce nouvel algorithme. Ensuite, dans la section 3, nous l'adaptions au cas d'une matrice symplectique  $W$  afin de déterminer les projecteurs  $\mathbb{P}_0$ ,  $\mathbb{P}_1$  et  $\mathbb{P}_\infty$  associés respectivement aux valeurs propres situées à l'intérieur du cercle unité, sur le cercle unité et à l'extérieur du cercle unité. Lorsque  $\mathbb{P}_0 = \mathbb{P}_\infty = 0$  et  $W$  n'a pas les valeurs propres  $-1$  et  $1$ , nous proposons dans la section 4 une étude basée encore sur la dichotomie spectrale par rapport à des cercles, mais appliquée cette fois-ci à  $(W + I)^{-1}(W - I)$ . Cela permet l'analyse de la stabilité forte. L'analyse est basée sur le regroupement des projecteurs  $\mathbb{P}_r$  et  $\mathbb{P}_v$  associés aux  $r$ - et  $v$ -valeurs propres et théorème 7 du chapitre 1.

## 2.2 Dichotomie spectrale d'une matrice par rapport à un cercle

Soit  $A \in \mathbb{C}^{N \times N}$  une matrice n'ayant pas de valeur propre sur le cercle unité  $C(0, 1)$ . Le projecteur spectral sur le s-e-i associé aux valeurs propres intérieures à  $C(0, 1)$  est donné par (voir par exemple [15, p. 39])

$$\mathbb{P} = \frac{1}{2i\pi} \int_{C(0,1)} (zI - A)^{-1} dz = \frac{1}{2\pi} \int_0^{2\pi} (I - e^{-i\theta} A)^{-1} d\theta. \quad (2.1)$$

---

<sup>1</sup>L'algorithme permet de calculer la matrice  $H$  et pas seulement  $\|H\|_2$  comme dans les références citées précédemment, ce qui permet, en particulier, d'analyser la stabilité au sens de Lyapunov comme on le verra dans la section suivante.

Son calcul sera accompagné par celui de la matrice  $H$  définie par

$$H = \frac{1}{2\pi} \int_0^{2\pi} \left( (I - e^{-i\theta} A)^{-1} \right)^* (I - e^{-i\theta} A)^{-1} d\theta. \quad (2.2)$$

Le lien entre  $\mathbb{P}$  et  $H$  peut être brièvement expliqué par le fait que le couple  $(\mathbb{P}, H)$  est l'unique solution du système de Lyapunov généralisé [8]

$$\begin{cases} H - A^* H A = \mathbb{P}^* \mathbb{P} - (I - \mathbb{P})^* (I - \mathbb{P}) \\ H = H^* > 0 \\ \mathbb{P} A = A \mathbb{P} \\ \mathbb{P}^2 = \mathbb{P} \\ H \mathbb{P} = (H \mathbb{P})^*. \end{cases}$$

De plus, pour tout vecteur  $x$  et tout entier  $k$ , on a les estimations [24] :

$$\begin{aligned} \|A^k \mathbb{P} x\|_2 &\leq \sqrt{\|H\|_2 \|H^{-1}\|_2} \left( 1 - \frac{1}{\|H\|_2} \right)^{\frac{k}{2}} \|x\|_2, \\ \|A^k x\|_2 &\geq \frac{1}{\sqrt{\|H\|_2 \|H^{-1}\|_2}} \left( 1 + \frac{1}{\|H\|_2} \right)^{\frac{k}{2}} \|(I - \mathbb{P})x\|_2, \end{aligned}$$

qui montrent, entre autres, l'importance de la quantité  $\|H\|_2$  sur la décroissance asymptotique vers 0 (ou la croissance vers  $+\infty$ ) des puissances de  $A$ .

Un algorithme de dichotomie spectrale permettant un calcul stable et itératif de  $\mathbb{P}$  et  $H$  a été proposé dans [11]. Dans ce qui suit, nous allons rappeler les étapes les plus importantes de cet algorithme, pour ensuite, en proposer une variante tout aussi robuste mais économique en nombre d'opérations arithmétiques.

La fonction  $2\pi$ -périodique  $\theta \rightarrow (I_N - e^{-i\theta} A)^{-1}$  se décompose en série de Fourier :

$$(I - e^{-i\theta} A)^{-1} = \sum_{k=-\infty}^{+\infty} Z_k e^{ik\theta} \quad (2.3)$$

avec

$$Z_k = \frac{1}{2\pi} \int_0^{2\pi} (I_N - e^{-i\theta} A)^{-1} e^{-ik\theta} d\theta. \quad (2.4)$$

Alors on vérifie que

$$\begin{cases} Z_0 = \mathbb{P}, \\ H = \sum_{k=-\infty}^{+\infty} Z_k^* Z_k. \end{cases}$$



Pour avoir  $H$ , il faut donc, a priori, calculer toutes les matrices  $Z_k$ . On verra que ce n'est pas le cas. Les relations (2.3) et (2.4) s'écrivent sous la forme d'un système linéaire infini dont les inconnues sont les matrices  $Z_k$  (coefficients de Fourier) :

$$\begin{cases} Z_k - AZ_{k+1} = 0 & \text{si } k \neq 0, \\ Z_0 - AZ_1 = I_N. \end{cases} \quad (2.5)$$

A priori, la résolution de ce système semble compliquée, mais, en fait, les coefficients de Fourier  $Z_k$  n'ont pas tous un rôle significatif car d'une part on a [25] :

$$\|Z_k\|_2 \leq \sqrt{\|H\|_2 \|H^{-1}\|_2} \sqrt{\|H\|_2} \left(1 - \frac{1}{\|H\|_2}\right)^{-\frac{k}{2}} \quad \text{si } k \leq 0,$$

$$\|Z_k\|_2 \leq \sqrt{\|H\|_2 \|H^{-1}\|_2} \sqrt{\|H\|_2 + 1} \left(1 + \frac{1}{\|H\|_2}\right)^{-\frac{k}{2}} \quad \text{si } k \geq 1,$$

qui montre que les coefficients  $Z_k$  tendent rapidement vers 0 pour  $|k|$  grand, et d'autre part, comme il est montré dans [11], si on pose

$$Z_k^{(s)} = \sum_{l=-\infty}^{+\infty} Z_{k+ls}, \quad k \in \mathbb{Z},$$

où  $s$  est un entier donné, alors la suite  $\left(Z_k^{(s)}\right)_{k \in \mathbb{Z}}$  est  $s$ -périodique, i.e.  $Z_{k+s}^{(s)} = Z_k^{(s)}$ , et satisfait

$$\lim_{s \rightarrow +\infty} Z_k^{(s)} = Z_k \quad \forall k \in \mathbb{Z}.$$

Ceci justifie le fait qu'une approximation des coefficients de Fourier peut-être obtenue en résolvant, pour  $s$  grand, le système linéaire fini (et cyclique) d'inconnues  $Z_0^{(s)}, Z_1^{(s)}, \dots, Z_s^{(s)} = Z_0^{(s)}$

$$\begin{cases} Z_0^{(s)} - AZ_1^{(s)} = I \\ Z_k^{(s)} - AZ_{k+1}^{(s)} = 0 \quad 1 \leq k \leq s-1. \end{cases} \quad (2.6)$$

On montre aisément que ce système a une solution unique si et seulement si la matrice  $A$  n'a pas de valeur propre sur le cercle  $C(0, 1)$ , ce qui est notre hypothèse de départ.

Maintenant on considère des entiers  $s$  "grands" de la forme  $s = 2^{j+1}$ . Alors des approximations de  $\mathbb{P}$  et de  $H$  peuvent être obtenues de manière itérative grâce au théorème suivant [11].

**Théorème 8** *Posons  $B_0 = I$ ,  $A_0 = -A$  et*

$$\Delta_j = B_0 Z_{2^j}^{(2^{j+1})} + A_0 Z_1^{(2^{j+1})}, \quad (2.7)$$

$$\nabla_j = B_0 Z_{2^{j+1}}^{(2^{j+1})} + A_0 Z_{2^{j+1}}^{(2^{j+1})}, \quad (2.8)$$

$$H_j = \sum_{k=1}^{2^{j+1}} \left( Z_k^{(2^{j+1})} \right)^* Z_k^{(2^{j+1})}. \quad (2.9)$$

*Alors*

$$H_j = \Delta_j^* H_{j-1} \Delta_j + \nabla_j^* H_{j-1} \nabla_j. \quad (2.10)$$

*De plus*

$$\begin{aligned} \lim_{j \rightarrow \infty} Z_0^{(2^{j+1})} &\equiv \lim_{j \rightarrow \infty} Z_{2^{j+1}}^{(2^{j+1})} = \mathbb{P}, \\ \lim_{j \rightarrow \infty} Z_1^{(2^{j+1})} &= \mathbb{P} - I, \\ \lim_{j \rightarrow \infty} H_j &= H. \end{aligned}$$

Ce théorème montre en particulier que :

- l'approximation de  $\mathbb{P}$  nécessite seulement le calcul de  $Z_0^{(2^{j+1})} = Z_{2^{j+1}}^{(2^{j+1})}$  pour  $j$  grand,
- l'approximation de  $H$  nécessite seulement le calcul de  $Z_1^{(2^{j+1})}$ ,  $Z_{2^j}^{(2^{j+1})}$ ,  $Z_{2^{j+1}}^{(2^{j+1})}$  et  $Z_{2^{j+1}}^{(2^{j+1})}$  pour  $j$  grand.

Afin de déterminer les matrices  $Z_1^{(2^{j+1})}$ ,  $Z_{2^j}^{(2^{j+1})}$ ,  $Z_{2^{j+1}}^{(2^{j+1})}$  et  $Z_{2^{j+1}}^{(2^{j+1})}$ , il est procédé dans [11] de la manière suivante : ces matrices satisfont le système

$$\begin{pmatrix} B_0 & A_0 & 0 & 0 \\ A_j & 0 & B_j & 0 \\ 0 & B_j & 0 & A_j \\ 0 & 0 & A_0 & B_0 \end{pmatrix} \begin{pmatrix} Z_{2^j}^{(2^{j+1})} \\ Z_{2^{j+1}}^{(2^{j+1})} \\ Z_1^{(2^{j+1})} \\ Z_{2^{j+1}}^{(2^{j+1})} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ I \end{pmatrix} \quad (2.11)$$

où les matrices  $A_j$  et  $B_j$  sont construites itérativement à partir des matrices  $B_0$  et  $A_0$  à l'aide d'une factorisation  $QR$  obtenue par des transformations de Householder sous la forme

$$\begin{pmatrix} B_0 & A_0 & 0 & 0 \\ A_{j-1} & 0 & B_{j-1} & 0 \\ 0 & B_{j-1} & 0 & A_{j-1} \end{pmatrix} = Q^{(j-1)} \begin{pmatrix} R_{11}^{(j-1)} & R_{12}^{(j-1)} & R_{13}^{(j-1)} & R_{14}^{(j-1)} \\ 0 & R_{22}^{(j-1)} & R_{23}^{(j-1)} & R_{24}^{(j-1)} \\ 0 & 0 & B_j & A_j \end{pmatrix} \quad (2.12)$$

où  $Q^{(j-1)}$  est unitaire et les  $R_{kl}^{(j-1)}$  sont triangulaires supérieures. Cela permet une mise-à-jour des  $A_j$  et  $B_j$  et de résoudre par la suite le système (2.11).

Les étapes importantes de l'algorithme ainsi obtenu sont résumées ci-dessous (voir [11, Algorithm 1]).

**Algorithme 1** (*Dichotomie spectrale de  $A$  par le cercle unité*)

1. **Initialisation** :  $B_0 = I, A_0 = -A$  :

- Calculer  $Z_1^{(2)}$  et  $Z_2^{(2)}$  solutions du système

$$\begin{pmatrix} B_0 & A_0 \\ A_0 & B_0 \end{pmatrix} \begin{pmatrix} Z_1^{(2)} \\ Z_2^{(2)} \end{pmatrix} = \begin{pmatrix} 0 \\ I \end{pmatrix}.$$

- Calculer  $H_0 = (Z_1^{(2)})^* Z_1^{(2)} + (Z_2^{(2)})^* Z_2^{(2)}$

2. **Itération** : calcul de  $Z_1^{(2j+1)}, Z_{2j}^{(2j+1)}, Z_{2j+1}^{(2j+1)}, Z_{2j+1}^{(2j+1)}$  et  $H_j$  pour  $j = 1, 2, \dots$

- Calculer la factorisation  $QR$

$$\begin{pmatrix} B_0 & A_0 & 0 & 0 \\ A_{j-1} & 0 & B_{j-1} & 0 \\ 0 & B_{j-1} & 0 & A_{j-1} \end{pmatrix} = Q^{(j-1)} \begin{pmatrix} R_{11}^{(j-1)} & R_{12}^{(j-1)} & R_{13}^{(j-1)} & R_{14}^{(j-1)} \\ 0 & R_{22}^{(j-1)} & R_{23}^{(j-1)} & R_{24}^{(j-1)} \\ 0 & 0 & B_j & A_j \end{pmatrix}.$$

- Calculer  $Z_1^{(2j+1)}$  et  $Z_{2j+1}^{(2j+1)}$  solutions du système

$$\begin{pmatrix} B_j & A_j \\ A_0 & B_0 \end{pmatrix} \begin{pmatrix} Z_1^{(2j+1)} \\ Z_{2j+1}^{(2j+1)} \end{pmatrix} = \begin{pmatrix} 0 \\ I \end{pmatrix}.$$

- Calculer  $Z_{2j}^{(2j+1)}, Z_{2j+1}^{(2j+1)}$  solutions du système

$$\begin{pmatrix} R_{11}^{(j-1)} & R_{12}^{(j-1)} \\ 0 & R_{22}^{(j-1)} \end{pmatrix} \begin{pmatrix} Z_{2j}^{(2j+1)} \\ Z_{2j+1}^{(2j+1)} \end{pmatrix} = - \begin{pmatrix} R_{13}^{(j-1)} & R_{14}^{(j-1)} \\ R_{23}^{(j-1)} & R_{24}^{(j-1)} \end{pmatrix} \begin{pmatrix} Z_1^{(2j+1)} \\ Z_{2j+1}^{(2j+1)} \end{pmatrix}$$

- Calculer

$$\Delta_j = B_0 Z_{2j}^{(2j+1)} + A_0 Z_1^{(2j+1)},$$

$$\nabla_j = B_0 Z_{2j+1}^{(2j+1)} + A_0 Z_{2j+1}^{(2j+1)},$$

$$H_j = \Delta_j^* H_{j-1} \Delta_j + \nabla_j^* H_{j-1} \nabla_j.$$

**Estimation du nombre d'opérations dans l'algorithme 1 :**

On estime maintenant le nombre d'opérations élémentaires dans l'algorithme 1. Les estimations tiennent compte uniquement des termes les plus importants, i.e. les termes en  $N^3$ .

- Initialisation : pour calculer  $Z_1^{(2)}$  et  $Z_2^{(2)}$ , on résout les systèmes linéaires par blocs  $(I - A_0^2)Z_2^{(2)} = I$  et  $Z_1^{(2)} = -A_0Z_2^{(2)}$ . Le premier nécessite  $2N^3$  opérations (calcul de  $A_0^2$ ), puis  $2/3N^3$  (factorisation de Gauss de  $I - A_0^2$ ), puis  $2N^3$  résolution de deux systèmes triangulaires par blocs). Les calculs de  $Z_1^{(2)}$  et de  $H_0$  nécessitent respectivement  $2N^3 + 2N^3 = 4N^3$  opérations.  
La phase initialisation requiert donc  $26/3N^3$  opérations.
- Iteration  $j$  :
  - La factorisation  $QR$  par des transformations de Householder nécessite ici  $54N^3$  opérations [13].
  - Le calcul de  $Z_1^{(2^{j+1})}$  et  $Z_{2^{j+1}}^{(2^{j+1})}$  nécessite  $31/3N^3$  opérations :  $2/3(2N)^3$  pour la factorisation LU de  $\begin{pmatrix} B_j & A_j \\ A_0 & B_0 \end{pmatrix}$ , suivie de la résolution de deux systèmes triangulaires par blocs, et en tenant compte de la spécificité du second membre  $\begin{pmatrix} 0 \\ I \end{pmatrix}$ , ces systèmes nécessitent  $N^3$  pour L et  $4N^3$  pour U.
  - Le calcul de  $Z_{2^j}^{(2^{j+1})}$  et  $Z_{2^{j+1}}^{(2^{j+1})}$  nécessite  $12N^3 : 8N^3$  pour calculer  $\begin{pmatrix} R_{13}^{(j-1)} & R_{14}^{(j-1)} \\ R_{23}^{(j-1)} & R_{24}^{(j-1)} \end{pmatrix} \begin{pmatrix} Z_1^{(2^{j+1})} \\ Z_{2^{j+1}}^{(2^{j+1})} \end{pmatrix}$  et  $4N^3$  pour résoudre le système bloc-triangular à matrice  $\begin{pmatrix} R_{11}^{(j-1)} & R_{12}^{(j-1)} \\ 0 & R_{22}^{(j-1)} \end{pmatrix}$ .
  - Enfin le calcul de  $\Delta_j$ ,  $\nabla_j$  et  $H_j$  nécessite respectivement  $2N^3$ ,  $2N^3$  et  $4N^3$  opérations.

Au total, une itération nécessite de l'ordre de  $253/3N^3$  opérations. Le coût des opérations est résumé dans le tableau 2.1.

**2.2.1 Une variante plus économique de l'algorithme 1**

La variante de l'algorithme 1 que nous proposons supprime essentiellement la factorisation  $QR$  et le calcul de  $Z_{2^j}^{(2^{j+1})}$  et  $Z_{2^{j+1}}^{(2^{j+1})}$  de l'algorithme 1. Elle se justifie facilement de la façon suivante : la résolution du système (2.6), pour  $s = 2^{j+1}$ , donne  $2^{j+1}$  solutions d'expressions

$$Z_k^{(2^{j+1})} = A^{2^{j+1}-k} [I_N - A^{2^{j+1}}]^{-1}, \quad k = 0, 1, \dots, 2^{j+1} - 1. \quad (2.13)$$

---



---

<u>Initialisation</u>		
Calcul de $Z_1^{(2)}, Z_2^{(2)}, H_0$		$26/3N^3$
<u>Une itération <math>j</math></u>		
factorisation $QR$		$54N^3$
Calcul de $Z_1^{2(j+1)}, Z_{2^{j+1}}^{2(j+1)}$		$31/3N^3$
Calcul de $Z_{2^j}^{2(j+1)}, Z_{2^{j+1}}^{2(j+1)}$		$12N^3$
Calcul de $\Delta_j$		$2N^3$
Calcul de $\nabla_j$		$2N^3$
Calcul de $H_j$		$4N^3$
<hr/>		
Total	$\frac{26}{3}N^3 + \frac{253}{3}jN^3$	

---



---

TAB. 2.1 – Coût des opérations de l'algorithme 1

Evidemment le calcul analytique de ces expressions n'est pas stable. Il est préférable de les calculer de manière itérative. Pour cela, posons pour  $j = 0, 1, \dots$

$$K_{j+1} = A^{2^j} [I_N - A^{2^j}] [I_N - A^{2^{j+1}}]^{-1}, \quad (2.14)$$

$$L_{j+1} = [I_N - A^{2^j}] [I_N - A^{2^{j+1}}]^{-1}. \quad (2.15)$$

La proposition suivante donne le lien entre  $Z_k^{(2^{j+1})}$  et  $Z_k^{(2^j)}$ .

**Proposition 3** *Pour  $j = 0, 1, \dots$  et  $k = 0, 1, \dots, 2^j$ , on a*

$$Z_k^{(2^{j+1})} = Z_k^{(2^j)} K_{j+1}, \quad (2.16)$$

$$Z_{2^j+k}^{(2^{j+1})} = Z_k^{(2^j)} L_{j+1}, \quad (2.17)$$

$$H_{j+1} = K_{j+1}^* H_j K_{j+1} + L_{j+1}^* H_j L_{j+1}. \quad (2.18)$$

### Démonstration

On a

$$\begin{aligned}
Z_k^{(2^{j+1})} &= A^{2^{j+1}-k} [I_N - A^{2^{j+1}}]^{-1} \\
&= A^{2^j-k} [I_N - A^{2^j}]^{-1} A^{2^j} [I_N - A^{2^j}] [I_N - A^{2^{j+1}}]^{-1} \\
&= Z_k^{(2^j)} K_{j+1},
\end{aligned} \quad (2.19)$$

et

$$\begin{aligned}
Z_{2^j+k}^{(2^{j+1})} &= A^{2^{j+1}-(2^j+k)} [I_N - A^{2^j}]^{-1} [I_N - A^{2^j}] [I_N - A^{2^{j+1}}]^{-1} \\
&= A^{2^j-k_0} [I_N - A^{2^j}]^{-1} [I_N - A^{2^j}] [I_N - A^{2^{j+1}}]^{-1} \\
&= Z_k^{(2^j)} L_{j+1}.
\end{aligned} \tag{2.20}$$

D'autre part

$$\begin{aligned}
H_{j+1} &= \sum_{k=1}^{2^{j+1}} (Z_k^{(2^{j+1})})^* Z_k^{(2^{j+1})} \\
&= \sum_{k=1}^{2^j} (Z_k^{(2^{j+1})})^* Z_k^{(2^{j+1})} + \sum_{k=2^j+1}^{2^{j+1}} (Z_k^{(2^{j+1})})^* Z_k^{(2^{j+1})} \\
&= \sum_{k=1}^{2^j} K_{j+1}^* (Z_k^{(2^j)})^* Z_k^{(2^j)} K_{j+1} + \sum_{k=1}^{2^j} L_{j+1} (Z_k^{(2^j)})^* Z_k^{(2^j)} L_{j+1} \\
&= K_{j+1}^* \left( \sum_{k=1}^{2^j} (Z_k^{(2^j)})^* Z_k^{(2^j)} \right) K_{j+1} + L_{j+1} \left( \sum_{k=1}^{2^j} L_{j+1} (Z_k^{(2^j)})^* Z_k^{(2^j)} \right) L_{j+1} \\
&= K_{j+1}^* H_j K_{j+1} + L_{j+1}^* H_j L_{j+1}.
\end{aligned}$$

□

La proposition suivante permet de calculer les matrices  $K_{j+1}$  et  $L_{j+1}$  de manière itérative.

**Proposition 4** *Pour  $j = 0, 1, \dots$  on a*

$$\begin{pmatrix} B_j & A_j \\ A_j & B_j \end{pmatrix} \cdot \begin{pmatrix} K_{j+1} \\ L_{j+1} \end{pmatrix} = \begin{pmatrix} 0 \\ I_N \end{pmatrix} \tag{2.21}$$

avec

$$A_j = -AZ_1^{(2^j)}, \quad B_j = Z_{2^j}^{(2^j)}.$$

**Démonstration**

La deuxième ligne du système (2.6) écrite pour  $k = 2^j$  donne

$$Z_{2^j}^{(2^{j+1})} - AZ_{2^j+1}^{(2^{j+1})} = 0$$

ou, d'après la proposition 3

$$Z_{2^j}^{(2^j)} K_{j+1} - AZ_1^{(2^j)} L_{j+1} = 0$$

ou encore

$$B_j K_{j+1} + A_j L_{j+1} = 0. \quad (2.22)$$

De même, la première ligne s'écrit, sachant que  $Z_0^{(2^j)} = Z_{2^j}^{(2^j)}$

$$Z_{2^j}^{(2^j)} L_{j+1} - A Z_1^{(2^j)} K_{j+1} = I_N$$

ou encore

$$B_j L_{j+1} + A_j K_{j+1} = I_N. \quad (2.23)$$

□

Connaissant  $K_{j+1}$  et  $L_{j+1}$ , on détermine les matrices  $Z_1^{(2^{j+1})}$  et  $Z_{2^{j+1}}^{(2^{j+1})}$  à partir de  $Z_1^{(2^j)}$  et  $Z_{2^j}^{(2^j)}$  par application de la proposition 3, puis  $K_{j+2}$  et  $L_{j+2}$  comme solutions du système matriciel (2.21).

**Remarque 1** *Tous les résultats obtenus s'appliquent au cas où  $A$  n'a pas de valeurs propres sur le cercle  $C(0, r)$  de centre 0 et de rayon  $r$ . Il suffit en effet de remplacer  $A$  par  $\frac{A}{r}$ . C'est ce que nous supposons par la suite. La nouvelle variante de l'algorithme 1 est donnée ci-dessous (Algorithme 2)*

**Algorithme 2** (Dichotomie spectrale de  $A$  par le cercle  $C(0, r)$ )

(1) **Initialisation**

- (a)  $A_0 = -\frac{A}{r}$
- (b) Résoudre

$$\begin{pmatrix} A_0 & I_N \\ I_N & A_0 \end{pmatrix} \begin{pmatrix} K_1 \\ L_1 \end{pmatrix} = \begin{pmatrix} 0 \\ I_N \end{pmatrix}$$

- (c) Poser  $Z_1^{(2)} = K_1$ ,  $Z_2^{(2)} = L_1$  et calculer  $H_1 = (Z_1^{(2)})^* Z_1^{(2)} + (Z_2^{(2)})^* Z_2^{(2)}$ .

(2) **Itération** : pour  $j = 1, 2, \dots$

- (a) Poser

$$A_j = A_0 Z_1^{(2^j)}, \quad B_j = Z_{2^j}^{(2^j)}$$

- (b) Résoudre

$$\begin{pmatrix} B_j & A_j \\ A_j & B_j \end{pmatrix} \begin{pmatrix} K_{j+1} \\ L_{j+1} \end{pmatrix} = \begin{pmatrix} 0 \\ I_N \end{pmatrix}$$

(c) *Calculer*

$$\begin{aligned} Z_1^{(2^{j+1})} &= Z_1^{(2^j)} K_{j+1}, \\ Z_{2^{j+1}}^{(2^{j+1})} &= Z_{2^j}^{(2^j)} L_{j+1}, \\ H_{j+1} &= K_{j+1}^* H_j K_{j+1} + L_{j+1}^* H_j L_{j+1}. \end{aligned}$$

**Estimation du nombre d'opérations dans l'algorithme 2 :** Le calcul se fait de manière similaire à l'algorithme 1. Le coût des opérations est résumé dans le tableau 2.2.

<hr/>		
<u>Initialisation</u>		
Calcul de $K_1, L_1, H_1$		$26/3N^3$
<u>Une itération <math>j</math></u>		
Calcul de $A_j$		$2N^3$
Calcul de $K_{j+1}, L_{j+1}$		$31/3N^3$
Calcul de $Z_1^{(2^{j+1})}$		$2N^3$
Calcul de $Z_{2^{j+1}}^{(2^{j+1})}$		$2N^3$
Calcul de $H_{j+1}$		$4N^3$
<hr/>		
Total		$\frac{26}{3}N^3 + \frac{61}{3}jN^3$

TAB. 2.2 – Coût des opérations de l'algorithme 2

## Remarque 2

1. Le nombre d'opérations de la phase d'initialisation des algorithmes 1 et 2 est le même. En revanche une itération de l'algorithme 1 coûte plus de quatre fois plus cher qu'une itération de l'algorithme 2. L'algorithme 2 représente donc un gain non négligeable par rapport à l'algorithme 1.
2. En général, une dizaine d'itérations suffisent à l'algorithme 2 (et à l'algorithme 1) pour construire de très bonnes approximations du projecteur  $\mathbb{P}$  et de la matrice  $H$ .



### 2.2.2 Portrait spectral

Le portrait spectral de la matrice  $A$  est le graphe de la fonction

$$r \longmapsto f(r) = \|H(r)\| \quad (2.24)$$

où

$$H(r) = \frac{1}{2\pi} \int_0^{2\pi} \left( I - e^{i\theta} \frac{A^*}{r} \right)^{-1} \left( I - e^{-i\theta} \frac{A}{r} \right)^{-1} d\theta.$$

Le projecteur spectral sur le s-e-i de  $A$  associé aux valeurs propres  $\lambda$  intérieures au cercle  $C(0, r)$  est

$$\mathbb{P}(r) = \frac{1}{2i\pi} \int_{C(0,r)} (zI - A)^{-1} dz = \frac{1}{2\pi} \int_0^{2\pi} \left( I - e^{-i\theta} \frac{A}{r} \right)^{-1} d\theta.$$

Lorsque  $A$  a une valeur propre  $\lambda$  sur le cercle  $C(0, r)$ , alors  $f(r) \rightarrow \infty$ , i.e. le graphe de  $f$  a une asymptote d'équation  $|\lambda| = r$ . Il a été montré dans [8, Sec. 13] que la fonction  $f$  est convexe dans chaque intervalle où elle est définie. Ces intervalles correspondent aux régions d'absence de valeurs propres (la trace du projecteur y reste constante).

Le portrait spectral est un autre regard sur le pseudo-spectre [27] d'une matrice. Ce dernier permet la détermination de voisinages définis par les valeurs propres de toutes les perturbations de la matrice dans une région donnée. Plusieurs techniques permettent de le calculer, voir par exemple [27, 17, 2]. Le portrait spectral permet une stratification spectrale, i.e., la construction de régions (des bandes dans le cas qui nous intéresse) qui partagent le spectre. Une illustration en est donnée dans l'exemple suivant.

**Exemple 1** On calcule, par la méthode développée dans [22], le sous-espace invariant associé aux 10 plus petites valeurs propres de l'opérateur

$$-\nu \Delta u - \sum_{i=1}^3 \frac{\partial u}{\partial x_i}$$

discretisé par différences finies centrées, avec les conditions  $u = 0$  sur la frontière du cube unité  $\Omega = ]0, 1[^3$ . On prend  $\nu = 1/2$  et on utilise des grilles uniformes  $\Omega_l$  avec  $N_l = (2^l - 1)^3$ ,  $l = 2, 3, 4, 5$  points. Si on note par  $A_l$  l'opérateur discretisé sur la grille  $\Omega_l$ , l'algorithme développé dans [22] permet le calcul de la matrice  $Y_l$  dont les colonnes engendrent une base orthonormée du sous-espace invariant associé aux 10 plus petites valeurs propres de  $A_l$ . Notons par

$$B_l = Y_l^* A_l Y_l \quad \text{et} \quad \Delta_l = A_l Y_l - Y_l B_l.$$

La matrice  $B_l$  représente une approximation de la restriction de  $A$  sur la grille  $l$  et  $\Delta_l$  est le résidu associé à cette approximation. La figure 2.1 (à gauche) montre, pour la grille  $l = 5$ , le portrait spectral de  $B_l$ . Les asymptotes correspondent aux modules des valeurs propres rencontrées (multiplicité algébrique comprise). La trace du projecteur (i.e. le nombre de valeurs propres intérieures aux cercles  $C(0, r)$ ) est indiquée sur la figure. Remarquons que la valeur de la trace augmente à chaque traversée d'une asymptote. Le comportement du résidu  $\|\Delta_l\|$  au cours des itérations est indiqué à droite.

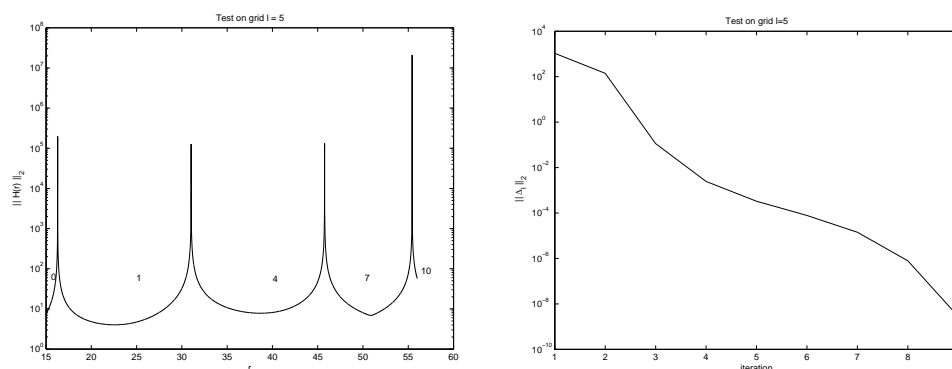


FIG. 2.1 – Portrait spectral (à gauche) et norme des résidus (à droite) sur la grille  $l = 5$ .

## 2.3 Application de l'algorithme aux matrices symplectiques

On se donne maintenant  $W \in \mathbb{R}^{2N \times 2N}$  et une matrice  $J \in \mathbb{R}^{2N \times 2N}$  inversible et anti-symétrique telles que  $W$  soit  $J$ -symplectique. Nous voulons déterminer les projecteurs  $\mathbb{P}_0$ ,  $\mathbb{P}_1$  et  $\mathbb{P}_\infty$  associés respectivement aux valeurs propres de module inférieur à 1, égal à 1 et supérieur à 1 à l'aide de l'algorithme 2. Pour cela, nous utilisons deux cercles  $C(0, r_0)$  et  $C(0, r_\infty)$  de même centre 0 et de rayons respectifs  $r_0$  et  $r_\infty$  tels que le premier contienne toutes les valeurs propres de  $W$  intérieures au cercle unité (et excluant les autres) et le second contienne toutes les valeurs propres de module inférieur ou égal à 1 (et excluant les autres). Alors, le projecteur spectral obtenu par l'algorithme 2 appliqué à  $W$  et au cercle  $C(0, r_0)$  sera égal à  $\mathbb{P}_0$  et celui obtenu avec  $C(0, r_\infty)$  sera égal à  $I_{2N} - \mathbb{P}_\infty$ . Nous obtenons ainsi  $\mathbb{P}_0$ ,  $\mathbb{P}_\infty$  et donc  $\mathbb{P}_1 = I_{2N} - \mathbb{P}_0 - \mathbb{P}_\infty$ .

Compte tenu du fait que les valeurs propres de la matrice symplectique  $W$  sont symétriques par rapport au cercle unité ( au sens de l'inversion ), il suffit de trouver  $r_0$  et de poser  $r_\infty = 1/r_0$ . Pour avoir  $r_0$ , il suffit, par exemple, de s'approcher du premier minimum de cette fonction situé juste à gauche de l'asymptote  $r = 1$ . Puisque la fonction  $r \rightarrow \|H(r)\|_2$  est convexe en dehors des asymptotes, ce minimum existe et peut être approché par différentes techniques. Une illustration est donnée dans l'exemple simple suivant.

**Exemple 2** On considère les matrices

$$B = \begin{pmatrix} \frac{4}{5} & -\frac{3}{5} & & & & \\ & \frac{4}{5} & & & & \\ & & 2 & & & \\ & & & 4 & & \\ & & & & 6 & 2 \\ & & & & -1 & 3 \end{pmatrix}, \quad W_0 = \begin{pmatrix} B & 0 \\ 0 & B^{-T} \end{pmatrix}, \quad J_0 = \begin{pmatrix} 0 & -I \\ I & 0 \end{pmatrix}.$$

La matrice  $W_0$  est  $J_0$ -symplectique. Son portrait spectral est tracé dans la figure 2.2.

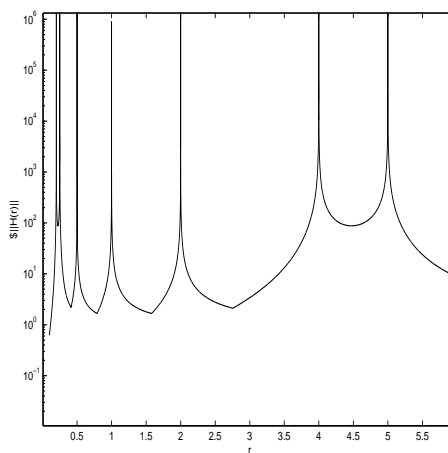


FIG. 2.2 – Portrait spectral de  $W_0$  (Exemple 2)

Une recherche simple dans la région du minimum local situé à gauche de  $r = 1$  donne une valeur de  $r_0$  de l'ordre de 0.9971. On vérifie que toutes les valeurs propres de  $W$  de module strictement inférieur à 1 sont à l'intérieur du cercle  $C(0, r_0)$ . L'application de l'algorithme 2 à  $W$  et au cercle  $C(0, r_0)$

donne

$$\mathbb{P}_0 = \begin{pmatrix} 0_8 & \\ & I_4 \end{pmatrix}, \quad \|H(r_0)\| = 170.67$$

( $0_8$  est la matrice nulle d'ordre 8 et  $I_4$  la matrice d'identité d'ordre 4).

En posant  $r_\infty = 1/r_0$  on détermine

$$I - \mathbb{P}_\infty = \begin{pmatrix} I_2 & & \\ & 0_4 & \\ & & I_6 \end{pmatrix}, \quad \|H(r_\infty)\| = 167.67.$$

Les valeurs de  $\|H(r_0)\|$  et  $\|H(r_\infty)\|$  montrent que la qualité numérique de  $\mathbb{P}_0$  et  $\mathbb{P}_\infty$  est bonne. On obtient

$$\mathbb{P}_\infty = \begin{pmatrix} 0_2 & & \\ & I_4 & \\ & & 0_6 \end{pmatrix} \quad \text{et} \quad \mathbb{P}_1 = \begin{pmatrix} I_2 & & \\ & 0_4 & \\ & & I_2 & \\ & & & 0_4 \end{pmatrix}.$$

## 2.4 Stabilité forte

Rappelons qu'une valeur propre  $\lambda$  de  $W$  de module 1 est dite  $r$ - (resp.  $v$ -) valeur propre si tout vecteur propre  $x$  associé à  $\lambda$  vérifie  $(S_0 x, x) > 0$  (resp.  $(S_0 x, x) < 0$ ), où  $S_0 = 1/2J(W - W^{-1}) \equiv 1/2(JW + (JW)^T)$ . Rappelons aussi que  $\pm 1$  sont des valeurs propres mêlées (cf. section 5 du chapitre 1). Enfin rappelons que si on désigne par  $\mathbb{P}_r$  le projecteur sur le s-e-i associé à toutes les  $r$ -valeurs propres de  $W$  et par  $\mathbb{P}_v$  celui associé aux  $v$ -valeurs propres, alors  $W$  est fortement stable si et seulement si  $\mathbb{P}_r + \mathbb{P}_v = \mathbb{P}_1 = I_{2N}$  (cf. théorème 7 du chapitre 1).

Remarquons que deux valeurs propres  $\lambda, \mu \neq \pm 1$  de  $W$  telles que  $\bar{\lambda}\mu = 1$  sont sur un même cercle, car  $\bar{\lambda}\mu = 1$  implique  $\frac{|\lambda-1|}{|\lambda+1|} = \frac{|\mu-1|}{|\mu+1|}$  qui est l'équation d'un cercle. Nous appliquons ici l'idée proposée initialement dans [5], et qui consiste à rassembler toutes les valeurs propres  $\lambda$  de  $W$  situées à l'intersection du cercle unité et du cercle d'équation  $\frac{|\lambda-1|}{|\lambda+1|} = cste$ . Voir figure 2.3. Nous savons d'après le théorème 1 du chapitre 1 que les vecteurs propres associés à des valeurs propres  $\lambda$  et  $\mu$  situées sur des cercles distincts ( $\frac{|\lambda-1|}{|\lambda+1|} \neq \frac{|\mu-1|}{|\mu+1|}$ ) sont  $J$ -orthogonaux. La conséquence de cette remarque simple nous permettra d'analyser la stabilité forte (détermination des projecteurs  $\mathbb{P}_r$  et  $\mathbb{P}_v$ ). Nous en résumons les idées principales :

Supposons que l'application de l'algorithme 2 à  $W$ , avec un choix adéquat de  $r_0$ , nous a fourni  $\mathbb{P}_0 = \mathbb{P}_\infty = 0$  et donc  $\mathbb{P}_1 = I$ .

Supposons par ailleurs que la matrice  $S_0$  est inversible et considérons la matrice

$$A = (W - I)(W + I)^{-1}$$

dont les valeurs propres sont de la forme

$$l_k = \frac{\lambda_k - 1}{\lambda_k + 1}$$

où  $\lambda_k$  désigne une valeur propre de  $W$  de module 1. Notons que les vecteurs propres (et les sous-espaces invariants) de  $A$  et de  $W$  sont les mêmes.

Soient  $a_1, a_2, \dots$  des nombres positifs qui entrelacent strictement les modules des valeurs propres  $|l_1|, |l_2|, \dots$ . Autrement dit,  $0 < a_k < |l_k| < a_{k+1}$  pour  $k = 1, \dots, m$ . Remarquons que puisque  $W$  est réelle, ses valeurs propres  $\lambda_k$  et par conséquent les  $l_k$  sont complexes conjuguées et donc  $m \leq N + 1$ . En pratique, il est possible de choisir par exemple  $a_1 = 0$  et  $a_{m+1}$  dans l'intervalle  $] \|A\|, \infty[$  où  $\|A\|$  désigne une norme matricielle quelconque de  $A$ . Les autres  $a_k$  peuvent être obtenus en examinant le portrait spectral de  $A$  comme dans l'exemple 3.

Nous savons, d'après la discussion ci-dessus, que les vecteurs propres (et les sous-espaces invariants) associés à des valeurs propres  $l_k$  et  $l_j$  dans des domaines différents sont  $J$ -orthogonaux. Ces régions peuvent être obtenues en appliquant l'algorithme 2 à la matrice  $A$ .

Notons par  $P_k$ ,  $k = 1, \dots, m + 1$  les projecteurs sur les s-e-i associés aux valeurs propres de  $A$  dans les cercles  $C(0, a_k)$ . Remarquons que  $P_1 = 0$  et  $P_{m+1} = I$ . D'autre part chaque matrice

$$Q_k = P_{k+1} - P_k \tag{2.25}$$

est un projecteur sur le sous-espace invariant de  $A$  (donc de  $W$ ) associé aux valeurs propres  $l = \frac{\lambda-1}{\lambda+1}$  de  $A$  situées dans la couronne

$$a_k < \frac{|\lambda - 1|}{|\lambda + 1|} < a_{k+1}. \tag{2.26}$$

Considérons maintenant la matrice  $S_k$  définie par

$$S_k = Q_k^T S_0 Q_k \equiv S_k^T. \tag{2.27}$$

L'idée est de rassembler toutes les valeurs propres  $\lambda$  de  $W$  pour lesquelles  $S_k$  est semi-définie positive (ou négative). Pour que  $W$  soit fortement stable, il ne doit pas y avoir de matrices  $S_k$  indéfinies.

Cette discussion permet de mieux préciser le théorème 7 de la manière suivante.

**Théorème 9** *La matrice  $W$  est fortement stable si et seulement si*

$$\mathbb{P}_0 = \mathbb{P}_\infty = 0 \quad \text{et} \quad \mathbb{P}_r + \mathbb{P}_v = \mathbb{P}_1 = I_{2N} \quad (2.28)$$

avec

$$\mathbb{P}_r = \sum_{S_k \geq 0} Q_k \quad \text{et} \quad \mathbb{P}_v = \sum_{S_k \leq 0} Q_k$$

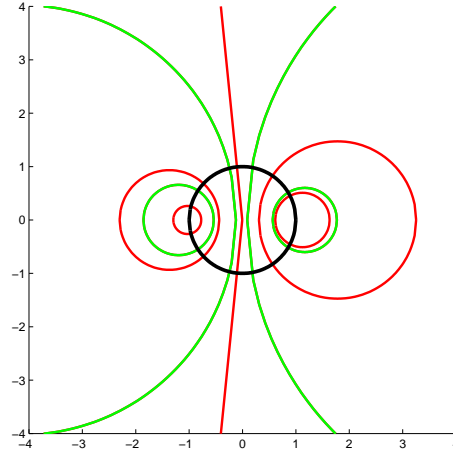


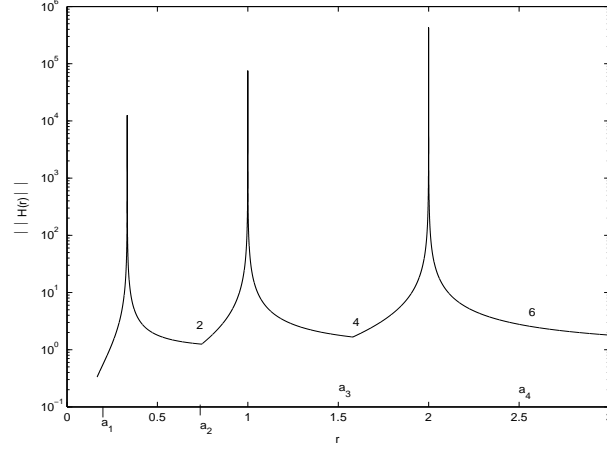
FIG. 2.3 – Partition du spectre de  $W$  par des cercles d'équations  $|\lambda - 1|/|\lambda + 1| = cste$ .

**Exemple 3** Cet exemple illustre le choix des nombres  $a_k$  évoqués dans la discussion ci-dessus.

On considère la matrice  $W$   $J$ -symplectique

$$W = \begin{pmatrix} \frac{4}{5} & 0 & 0 & 0 & \frac{3}{5} & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -\frac{3}{5} & 0 & \frac{4}{5} \\ -\frac{3}{5} & 0 & 0 & 0 & \frac{4}{5} & 0 \\ 0 & 0 & 0 & -\frac{4}{5} & 0 & -\frac{3}{5} \end{pmatrix}, \quad J = \begin{pmatrix} 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ -1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 & 0 & 0 \end{pmatrix}$$

dont les valeurs propres sont sur le cercle unité. Le portrait spectral de  $A = (W - I)(W + I)^{-1}$  est donné dans la figure 2.4.

FIG. 2.4 – Portrait spectral de  $A$  (Exemple 3)

Sur cette figure il y a trois asymptotes qui correspondent chacune à deux valeurs propres conjuguées de module 1 de  $W$  ( ou des valeurs propres de même couleur ). On peut déterminer des constantes  $a_k$ ,  $k = 1, 2, 3, 4$  situées entre les asymptotes et qui correspondent aux régions de petits critères de dichotomie( par exemple  $a_1 = 0$  et  $a_4 > \|W\|$  comme expliqué précédemment). D'autre part, il suffit de regarder la figure et choisir ces constantes. Par exemple on peut choisir  $a_1 = 0.1667$ ,  $a_2 = 0.6667$ ,  $a_3 = 1.5$  et  $a_4 = 2.5$ . En fait pour tout  $a_1$  appartenant à l'intervalle  $]0, 1/3[$ , on obtient le projecteur  $P_1 = 0$  en appliquant la dichotomie spectrale à  $A = (W - I)(W + I)^{-1}$  par rapport au cercle  $C(0, a_1)$ . De même, pour  $a_2 \in [1/3, 1[$ ,  $a_3 \in ]1, 2[$  et  $a_4 \in ]2, \infty[$ , on obtient les mêmes projecteurs  $P_2$ ,  $P_3$  et  $P_4 = I$  de traces respectives 2, 4 et 6. Ces traces ne varient pas dans les intervalles définis par les abscisses des différentes asymptotes. En revanche, la qualité numérique des projecteurs se dégrade en s'approchant des valeurs propres de  $A$ .

L'algorithme permettant de déterminer la stabilité forte d'une matrice  $W$   $J$ -symplectique est écrit ci-dessous de manière informelle (Algorithme 3).

### Algorithme 3

1. En utilisant l'algorithme 2 et un choix adéquat du paramètre  $r_0$ , déterminer les projecteurs  $\mathbb{P}_0$ ,  $\mathbb{P}_1$  et  $\mathbb{P}_\infty$ .
2. Si  $\mathbb{P}_0 \neq 0$  (ou  $\mathbb{P}_\infty \neq 0$ ), il n'y a pas stabilité forte. Sinon calculer

$$S_0 = (1/2)((JW) + (JW)^T).$$

Si  $S_0$  est singulière (mal conditionnée), il n'y a pas stabilité. Sinon calculer

$$A = (W - I)(W + I)^{-1}.$$

A partir du portrait spectral de  $A$ , déterminer des scalaires  $(a_k)_{1 \leq k \leq m+1}$  vérifiant (2.26).

3. Déterminer les projecteurs  $P_1 = 0$ ,  $P_2, \dots, P_m$ ,  $P_{m+1} = I$  par la dichotomie spectrale appliquée à  $A$  par rapport au cercle  $C(0, a_k)$ .
4. Pour  $k = 1, \dots, m$ , calculer

$$Q_k = P_{k+1} - P_k \quad \text{et} \quad S_k = Q_k^T S_0 Q_k.$$

Si  $S_k$  est indéfinie, alors il n'y a pas stabilité forte. Si toutes les matrices  $S_k$  sont semi-définies, alors il y a stabilité forte. Dans ce cas

$$\mathbb{P}_r = \sum_{S_k \geq 0} Q_k \quad \text{et} \quad \mathbb{P}_v = \sum_{S_k \leq 0} Q_k.$$

### Remarque 3

1. Comme nous l'avons déjà mentionné, on peut prendre  $a_1 = 0$  et  $a_{m+1} > \|A\|$ , et on a toujours  $P_0 = 0$  et  $P_{m+1} = I$ .
2. Dans l'algorithme 3, le calcul de chaque projecteur est accompagné du critère de dichotomie qui lui est associé. L'algorithme s'arrête si ce critère est trop grand. Lorsqu'il y a stabilité forte, tous les projecteurs doivent pouvoir être calculés avec un bon critère de dichotomie.
3. Pour définir les valeurs propres de couleurs rouge et verte, nous nous sommes restreints au cercle unité parce que nous sommes essentiellement intéressés par l'étude de la stabilité. Mais au delà des questions liées à la stabilité, il est tout à fait possible de définir ces valeurs propres dans un cadre plus général, i.e. non nécessairement sur le cercle unité. En effet, si  $\mathbb{P}_0, \mathbb{P}_\infty \neq 0$  et donc  $\mathbb{P}_1 \neq I$ , il suffit de répéter l'analyse précédente avec  $A = (\mathbb{P}_1 W - I)(\mathbb{P}_1 W + I)^{-1}$ . Une simple adaptation de l'algorithme 3 permet de savoir si le projecteur  $\mathbb{P}_1$  peut se décomposer sous la forme  $\mathbb{P}_1 = \mathbb{P}_r + \mathbb{P}_v$  avec  $\mathbb{P}_r = \sum_{S_k \geq 0} Q_k$  et  $\mathbb{P}_v = \sum_{S_k \leq 0} Q_k$ .

**Exemple 4** Considérons les matrices

$$W = \begin{pmatrix} B & 0 \\ 0 & B^{-T} \end{pmatrix} \quad \text{et} \quad J = \begin{pmatrix} 0 & -I \\ I & 0 \end{pmatrix}$$



où  $B \in \mathbb{R}^{N \times N}$  est diagonalisable et ses valeurs propres sur le cercle unité. On vérifie que la matrice  $W$  est  $J$ -symplectique et son spectre est sur le cercle unité. Ses vecteurs propres sont de la forme  $x = (u^T \ 0)^T$  ou  $x = (0 \ v^T)^T$  où  $u$  et  $v$  sont des vecteurs propres de  $B$  ou de  $B^{-T}$ . On vérifie alors que  $(S_0 x, x) = 0$ . Autrement dit, toutes les valeurs propres de  $W$  sont de type mêlé. Il n'y a donc pas stabilité forte. Remarquons que l'application de l'algorithme 3 à des matrices  $W$ , construites comme dans cet exemple, donne des matrices  $S_k$  indéfinies.

**Exemple 5** Dans cet exemple nous illustrons en détail la notion de stabilité forte. On considère les matrices

$$W(t) = \begin{pmatrix} A(s(t)) \cos \omega(t) & -(A(s(t)))^{-T} \sin \omega(t) \\ A(s(t)) \sin \omega(t) & (A(s(t)))^{-T} \cos \omega(t) \end{pmatrix}, \quad J = \begin{pmatrix} 0 & -I \\ I & 0 \end{pmatrix}$$

avec  $A(s) = \begin{pmatrix} 1-s^2 & -1 \\ s^2 & 1-s^2 \end{pmatrix}$ ,  $\omega(t) = \pi(\frac{1}{2} - \frac{1}{3} \sin 3t)$  et  $s(t) = 4 \sin(t)$ .

On vérifie facilement que  $W(t)$  est  $J$ -symplectique pour tout  $t$ . La fonction  $W(t)$  est  $2\pi$ -périodique et son spectre, pour  $t \in [0, 2\pi]$ , est donné en figure 2.5.

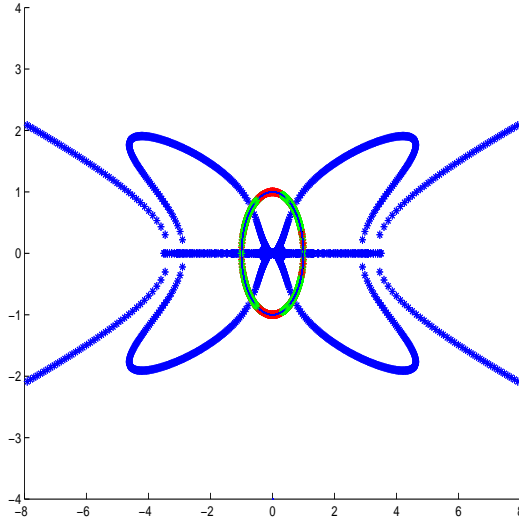


FIG. 2.5 – Distribution spectrale de  $W(t)$  lorsque  $t$  parcourt l'intervalle  $[0, 2\pi]$

Cette figure montre que les valeurs propres quittent le cercle unité en six endroits : en  $\pm 1$  et en  $\pm 0.673... \pm i0.739...$ , qui sont nécessairement des

valeurs propres mêlées. La valeur propre  $-0.673... + i0.739...$  est obtenue avec  $t_0 \approx 0.390017605$ . Voyons ce qui se passe autour de  $t_0$ .

a)  $t = 0.3896$

$$W = \begin{pmatrix} -1.0744 & -0.8213 & 0.1857 & 0.3276 \\ 1.8958 & -1.0744 & -0.1419 & 0.1857 \\ -0.7463 & -0.5705 & -0.2673 & -0.4716 \\ 1.3168 & -0.7463 & 0.2043 & -0.2673 \end{pmatrix}$$

L'algorithme 3 donne  $\mathbb{P}_0 = \mathbb{P}_\infty = 0$ ,  $\mathbb{P}_1 = I$ . Les portraits spectraux de  $W$  et de  $A = (W - I)(W + I)^{-1}$  sont donnés en figure 2.6.

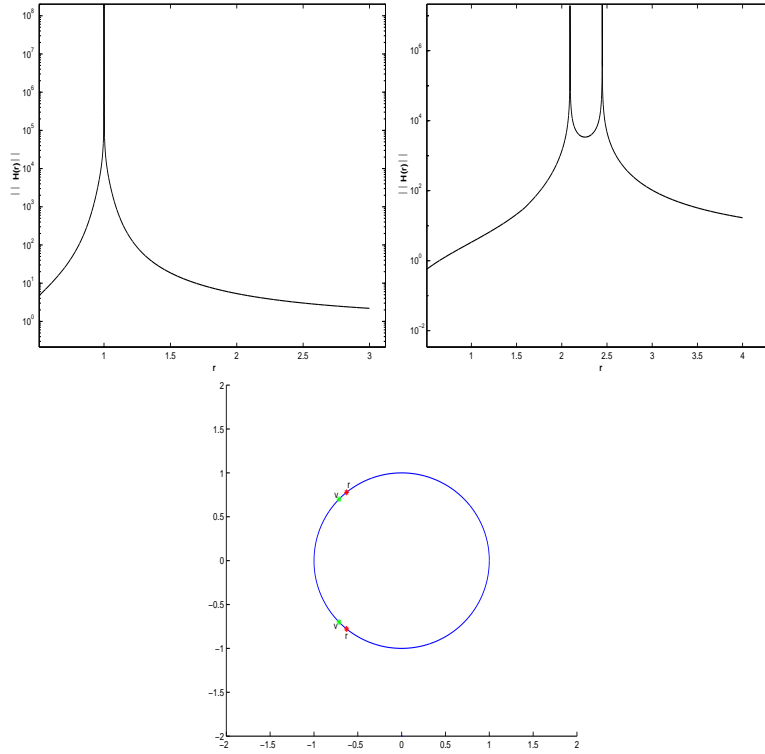


FIG. 2.6 – Portraits spectraux de  $W$  (en haut à gauche) et de  $A = (W - I)(W + I)^{-1}$  (en haut à droite) et la distribution spectrale de  $W$  à  $t = 0.3896$

Le portrait spectral de  $W$  montre que  $\mathbb{P}_0 = \mathbb{P}_\infty = 0$  et  $\mathbb{P}_1 = I$ . Le portrait spectral de  $A$  permet de déterminer des constantes  $a_1$ ,  $a_2$  et  $a_3$ , puis les projecteurs  $P_1 = 0$ ,  $P_2$  et  $P_3 = I$ , puis les projecteurs  $Q_1 = P_2$  et  $Q_2 = I - P_2$ , et ensuite les matrices

$$S_1 = Q_1^T S_0 Q_1 = \begin{pmatrix} 10.2636 & 1.5631 & -3.4267 & -0.7114 \\ 1.5631 & 7.3330 & -0.8454 & -5.4623 \\ -3.4267 & -0.8454 & 1.1589 & 0.4817 \\ -0.7114 & -5.4623 & 0.4817 & 4.0895 \end{pmatrix}$$

$$S_2 = Q_2^T S_0 Q_2 = \begin{pmatrix} -9.5173 & -1.9362 & 3.0232 & 1.8951 \\ -1.9362 & -6.5867 & 0.3326 & 5.0587 \\ 3.0232 & 0.3326 & -0.9732 & -0.3889 \\ 1.8951 & 5.0587 & -0.3889 & -3.9038 \end{pmatrix}.$$

dont les valeurs propres sont 0, 13.3181 et 9.5269 pour  $S_1$  et 0, -13.2030 et -7.7781 pour  $S_2$ . La matrices  $S_1$  et  $S_2$  sont donc semi-définies positives et semi-définie négative respectivement. On en déduit que  $W$  est fortement stable et

$$\mathbb{P}_r = Q_1 = \begin{pmatrix} 0.5000 & -3.5843 & -0.0000 & 2.7276 \\ 8.2733 & 0.5000 & -2.7276 & -0.0000 \\ 0.0000 & -10.9636 & 0.5000 & 8.2733 \\ 10.9636 & 0.0000 & -3.5843 & 0.5000 \end{pmatrix}$$

$$\mathbb{P}_v = Q_2 = \begin{pmatrix} 0.5000 & 3.5843 & 0.0000 & -2.7276 \\ -8.2733 & 0.5000 & 2.7276 & 0.0000 \\ -0.0000 & 10.9636 & 0.5000 & -8.2733 \\ -10.9636 & -0.0000 & 3.5843 & 0.5000 \end{pmatrix}.$$

Et on a  $\|\mathbb{P}_r^2 - \mathbb{P}_r\| = 2.6901e - 013$ ,  $\|\mathbb{P}_v^2 - \mathbb{P}_v\| = 2.6901e - 013$ ,  $\|\mathbb{P}_r \cdot \mathbb{P}_v\| = 2.6909e - 013$ ,  $\|\mathbb{P}_r \cdot W - W \mathbb{P}_r\| = 3.5178e - 012$  et  $\|\mathbb{P}_v W - W \mathbb{P}_v\| = 3.5184e - 012$ . Les deux projecteurs sont de trace 2. On a deux valeurs propres  $-0.6278 \pm 0.7783i$  de couleur rouge et deux valeurs propres  $-0.7139 \pm 0.7002i$  de couleur verte. La distribution spectrale est donnée en figure 2.6.

b)  $t = 0.39$

$$W = \begin{pmatrix} -1.0785 & -0.8216 & 0.1854 & 0.3267 \\ 1.9001 & -1.0785 & -0.1413 & 0.1854 \\ -0.7483 & -0.5701 & -0.2672 & -0.4708 \\ 1.3184 & -0.7483 & 0.2036 & -0.2672 \end{pmatrix}.$$

L'algorithme 3 donne  $\mathbb{P}_0 = \mathbb{P}_\infty = 0$ ,  $\mathbb{P}_1 = I$ . Les portraits spectraux de  $W$  et de  $A = (W - I)(W + I)^{-1}$  sont donnés en figure 2.7.

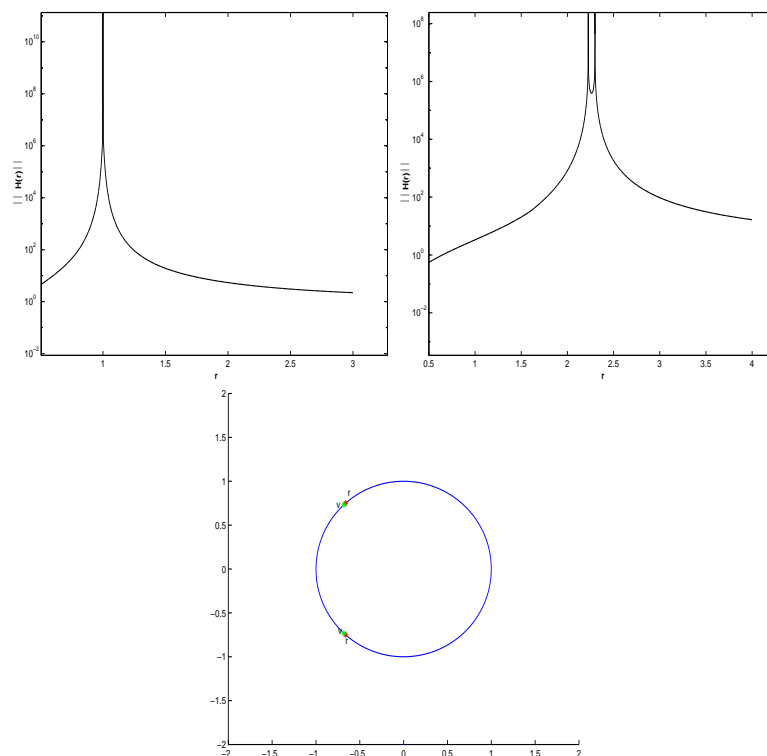


FIG. 2.7 – Portraits spectraux de  $W$  (en haut à gauche) et de  $A = (W - I)(W + I)^{-1}$  (en haut à droite) et la distribution spectrale de  $W$  à  $t = 0.39$

Les matrices  $S_k$  sont

$$S_1 = \begin{pmatrix} 48.5918 & 8.4029 & -15.9809 & -5.9051 \\ 8.4029 & 34.3108 & -3.0710 & -25.8898 \\ -15.9809 & -3.0710 & 5.2587 & 2.1748 \\ -5.9051 & -25.8898 & 2.1748 & 19.5397 \end{pmatrix}$$

$$S_2 = \begin{pmatrix} -47.8434 & -8.7770 & 15.5753 & 7.0906 \\ -8.7770 & -33.5625 & 2.5584 & 25.4842 \\ 15.5753 & 2.5584 & -5.0733 & -2.0821 \\ 7.0906 & 25.4842 & -2.0821 & -19.3542 \end{pmatrix}$$

avec les valeurs propres 0, 64.7756 et 42.9253 pour  $S_1$  et 0, -64.6609 et -41.1726 pour  $S_2$ . On en déduit que  $W$  est fortement stable, et on

a

$$\mathbb{P}_r = \begin{pmatrix} 0.5000 & -17.4891 & 0.0000 & 13.2415 \\ 40.4469 & 0.5000 & -13.2415 & -0.0000 \\ -0.0000 & -53.4405 & 0.5000 & 40.4469 \\ 53.4405 & 0.0000 & -17.4891 & 0.5000 \end{pmatrix}$$

$$\mathbb{P}_v = \begin{pmatrix} 0.5000 & 17.4891 & -0.0000 & -13.2415 \\ -40.4469 & 0.5000 & 13.2415 & 0.0000 \\ 0.0000 & 53.4405 & 0.5000 & -40.4469 \\ -53.4405 & -0.0000 & 17.4891 & 0.5000 \end{pmatrix}$$

avec  $\|\mathbb{P}_r^2 - \mathbb{P}_r\| = 1.5599e-011$ ,  $\|\mathbb{P}_v^2 - \mathbb{P}_v\| = 1.5599e-011$ ,  $\|\mathbb{P}_r \cdot \mathbb{P}_v\| = 1.5602e-011$ ,  $\|\mathbb{P}_r W - W \mathbb{P}_r\| = 1.2965e-009$  et  $\|\mathbb{P}_v W - W \mathbb{P}_v\| = 1.2965e-009$ . Les deux projecteurs sont de trace 2. La distribution spectrale est donnée en figure 2.7.

c)  $t = 0.390017605$ 

$$W = \begin{pmatrix} -1.0787 & -0.8216 & 0.1854 & 0.3266 \\ 1.9003 & -1.0787 & -0.1412 & 0.1854 \\ -0.7484 & -0.5701 & -0.2672 & -0.4708 \\ 1.3185 & -0.7484 & 0.2035 & -0.2672 \end{pmatrix}$$

L'algorithme 3 donne  $\mathbb{P}_0 = \mathbb{P}_\infty = 0$ ,  $\mathbb{P}_1 = I$ . Les portraits spectraux de  $W$  et de  $A = (W - I)(W + I)^{-1}$  sont donnés en figure 2.8.

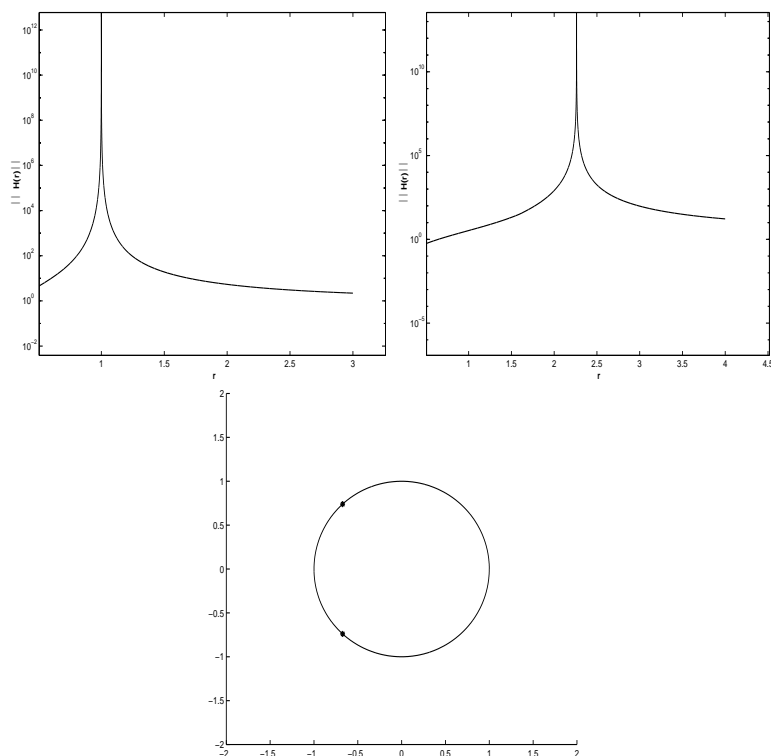


FIG. 2.8 – Portraits spectraux de  $W$  (en haut à gauche) et de  $A = (W - I)(W + I)^{-1}$  (en haut à droite) et la distribution spectrale de  $W$  à  $t = 0.390017605$

L'algorithme 3 construit une seule matrice

$$S_0 = \begin{pmatrix} 0.7484 & -0.3742 & -0.4057 & 1.1855 \\ -0.3742 & 0.7484 & -0.5126 & -0.4057 \\ -0.4057 & -0.5126 & 0.1854 & 0.0927 \\ 1.1855 & -0.4057 & 0.0927 & 0.1854 \end{pmatrix}$$

indéfinie. La matrice  $W$  n'est donc pas fortement stable. Aux instants après, les valeurs propres quittent le cercle unité comme le montre l'exemple suivant.

d)  $t = 0.39050$

$$W = \begin{pmatrix} -1.0836 & -0.8219 & 0.1851 & 0.3255 \\ 1.9055 & -1.0836 & -0.1404 & 0.1851 \\ -0.7509 & -0.5696 & -0.2671 & -0.4698 \\ 1.3204 & -0.7509 & 0.2026 & -0.2671 \end{pmatrix}$$

L'algorithme 3 donne  $\mathbb{P}_1 = 0$ ,  $\mathbb{P}_0 \neq 0$  et  $\mathbb{P}_\infty \neq 0$  avec

$$\|\mathbb{P}_1^2 - \mathbb{P}_1\| = 2.96e-012, \|\mathbb{P}_0^2 - \mathbb{P}_0\| = 3.31e-013, \|\mathbb{P}_\infty^2 - \mathbb{P}_\infty\| = 8.97e-014.$$

$$\|\mathbb{P}_0 W - W \mathbb{P}_0\| = 3.20e-012, \|\mathbb{P}_\infty W - W \mathbb{P}_\infty\| = 1.70e-012.$$

$W$  n'est pas stable. Les portraits spectraux de  $W$  et de  $A$ , et la distribution spectrale de  $W$  sont donnés en figure 2.9.

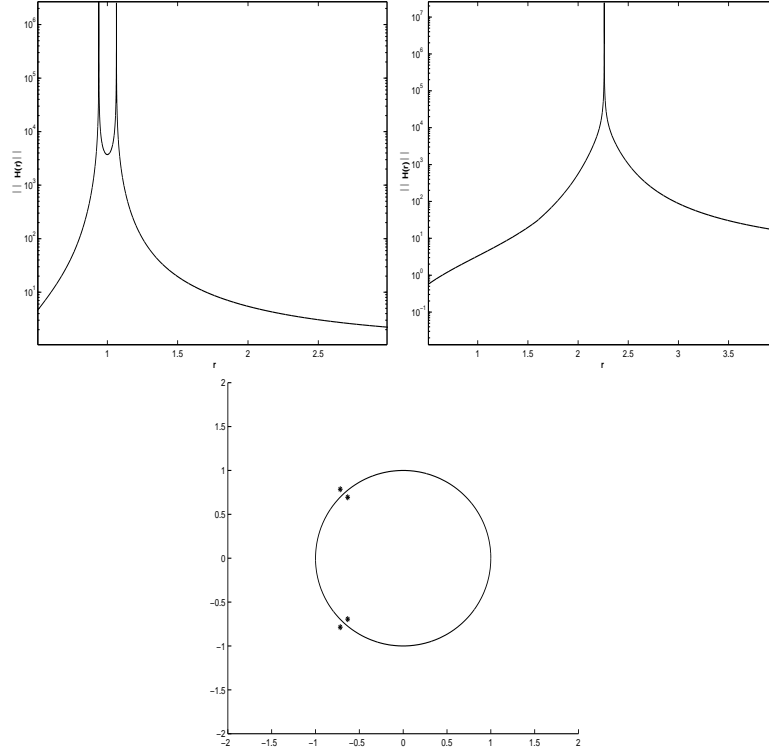


FIG. 2.9 – Portraits spectraux de  $W$  (en haut à gauche) et de  $A = (W - I)(W + I)^{-1}$  (en haut à droite) et la distribution spectrale de  $W$  à  $t = 0.3905$

## 2.5 Conclusion

Ce chapitre a porté sur une adaptation des méthodes de dichotomie spectrale aux matrices symplectiques. L'algorithme proposé (Algorithme 2) est

aussi robuste que celui proposé dans [11, Algorithm 1], mais il est quatre fois moins coûteux que ce dernier.

En examinant le portrait spectral, on détermine un paramètre  $r_0$  tel que le cercle  $C(0, r_0)$  contienne toutes les valeurs propres de module strictement inférieur à 1 d'une matrice symplectique  $W$ . Il faut deux applications de l'algorithme 2. Une avec  $W$  et  $C(0, r_0)$  et une autre avec  $W$  et  $C(0, r_\infty)$ ,  $r_\infty = 1/r_0$  pour finalement obtenir les trois projecteurs  $\mathbb{P}_0$ ,  $\mathbb{P}_\infty$  et  $\mathbb{P}_1$  sur les s-e-i associés respectivement aux valeurs propres intérieures, extérieures, et sur le cercle unité. Le calcul de ces trois projecteurs est accompagné du critère de dichotomie, une sorte de conditionnement, donnant une information qualitative sur les projecteurs calculés.

Dans le cas où  $\mathbb{P}_0 = \mathbb{P}_\infty = 0$  et donc  $\mathbb{P}_1 = I$ , l'algorithme 3 permet de savoir s'il est possible de décomposer le projecteur  $\mathbb{P}_1 = I$  en deux projecteurs supplémentaires  $\mathbb{P}_r$  et  $\mathbb{P}_v$  sur les s-e-i associés respectivement aux  $r-$  et  $v-$ valeurs propres. Cette décomposition est possible si et seulement si la matrice  $W$  est fortement stable (cf. théorème 7).

Au delà de la stabilité forte, l'algorithme 3 permet, en fait, de déterminer les projecteurs  $\mathbb{P}_r$  et  $\mathbb{P}_v$  indépendamment de la condition  $\mathbb{P}_0 = \mathbb{P}_\infty = 0$  (cf. remarque 3). Il suffit en effet d'appliquer l'algorithme 3 à

$$A = (\mathbb{P}_1 W - I)(\mathbb{P}_1 W + I)^{-1}.$$

De ce point de vue là, les algorithmes 2 et 3 permettent de réaliser une "quadrithomie" spectrale de  $W$  :  $\mathbb{P}_0, \mathbb{P}_\infty = 0, \mathbb{P}_1 = \mathbb{P}_r + \mathbb{P}_v$ .

Les algorithmes 2 et 3 sont robustes mais présentent un inconvénient : le coût global en nombre d'opérations pour déterminer les projecteurs  $\mathbb{P}_0, \mathbb{P}_\infty, \mathbb{P}_1, \mathbb{P}_r, \mathbb{P}_v$  reste élevé malgré les simplifications et les réductions du coût apportées par l'algorithme 2 (comparé à l'algorithme 1). Ceci est dû au choix de  $r_0$  qui peut-être coûteux (calcul d'un portrait spectral) et parfois difficile à déterminer. Les algorithmes proposés au chapitre 3 ne présentent pas ces inconvénients.





## Chapitre 3

# Une méthode de trichotomie spectrale de matrices symplectiques

### 3.1 Introduction

Ce chapitre est consacré à l'étude d'une nouvelle méthode, appelée par la suite "méthode de trichotomie", permettant de déterminer les sous-espaces invariants  $\mathcal{X}_0$ ,  $\mathcal{X}_1$  et  $\mathcal{X}_\infty$  d'une matrice  $J$ -symplectique  $W$  associés aux valeurs propres à l'intérieur, sur et à l'extérieur du cercle unité respectivement. L'objectif principal de ce calcul étant évidemment de déterminer si la matrice  $W$  est stable/fortement stable ou non.

Pour une analyse de stabilité, l'avantage de la méthode de trichotomie est qu'elle ne calcule pas les valeurs propres, mais des sous-espaces invariants. Or les sous-espaces invariants d'une matrice sont nettement moins sensibles aux perturbations que les valeurs propres [27, 8].

De ce point de vue, la méthode de trichotomie poursuit le même but que la méthode de dichotomie. Cependant, pour l'analyse de stabilité, l'utilisation de la méthode de dichotomie se heurte au problème du coût en nombre d'opérations à cause de la détermination du rayon  $r_0$  permettant de calculer de manière fiable les sous-espaces invariants  $\mathcal{X}_0$ ,  $\mathcal{X}_1$  et  $\mathcal{X}_\infty$  de  $W$ . La méthode de trichotomie proposée permet de s'affranchir de cette détermination.

Le principe de cette méthode repose sur le calcul de trois suites matri-

cielles  $(X_0^{(n)})_{n \geq 0}$ ,  $(X_1^{(n)})_{n \geq 0}$  et  $(X_\infty^{(n)})_{n \geq 0}$ . Chacune de ces suites doit permettre de déterminer une base de chacun des sous-espaces invariants  $\mathcal{X}_0$ ,  $\mathcal{X}_1$  et  $\mathcal{X}_\infty$ . Cependant ce processus n'est pas direct. En effet, seule la suite  $(X_1^{(n)})_{n \geq 0}$  permet de calculer une base de  $(X_1^{(n)})_{n \geq 0}$ . Les deux autres suites  $(X_0^{(n)})_{n \geq 0}$  et  $(X_\infty^{(n)})_{n \geq 0}$  convergent, en un certain sens qui sera défini, vers des bases des sous-espaces  $\mathcal{X}_0 + \mathcal{X}_1$  et  $\mathcal{X}_\infty + \mathcal{X}_1$  respectivement.

Afin de déterminer des bases de  $\mathcal{X}_0$  et  $\mathcal{X}_\infty$ , la méthode de trichotomie nécessite une étape que nous nommerons "étape de  $J$ -orthogonalisation". Elle repose sur le fait que  $(X_1^{(n)})_{n \geq 0}$  permet d'approcher  $\mathcal{X}_1$  et peut être utilisée pour éliminer dans les suites  $(X_0^{(n)})_{n \geq 0}$  et  $(X_\infty^{(n)})_{n \geq 0}$ , les composantes associées à  $\mathcal{X}_1$  et par conséquent d'en extraire des estimations de  $\mathcal{X}_0$  et  $\mathcal{X}_\infty$ .

Lorsque le spectre de  $W$  est situé sur le cercle unité, Godunov a proposé une méthode permettant de déterminer si  $W$  est stable/ fortement stable [6]. Cette approche n'utilise pas de dichotomie, elle ne fait intervenir que des calculs spectraux sur des matrices symétriques (auto-adjointes) pour lesquelles le calcul se fait avec une précision garantie [21]. Elle peut remplacer l'approche discutée dans la section 2 du chapitre 2.

L'étude de la méthode de trichotomie suit ces différentes étapes. Dans la section 2, on donne les définitions et certaines propriétés d'outils mathématiques utilisés dans les sections suivantes. Les sections 3, 4 et 5 sont consacrées respectivement à la convergence des suites  $(X_0^{(n)})_{n \geq 0}$ ,  $(X_1^{(n)})_{n \geq 0}$  et  $(X_\infty^{(n)})_{n \geq 0}$ , à l'étude de l'étape de  $J$ -orthogonalisation et à la qualité finale des estimations calculées par la méthode de trichotomie. La mise en œuvre de cette méthode est étudiée dans la section 6 (voir en particulier l'algorithme 4). L'approche proposée par Godunov permettant d'analyser la stabilité / stabilité forte d'une matrice est mentionnée dans la section 7. Le chapitre se termine par la section 8 où sont illustrées des applications numériques.

## 3.2 Notations et préliminaires techniques

### 3.2.1 Principales notations

Afin de simplifier la lecture de ce chapitre, toutes les notations utilisées sont regroupées dans cette section.

Soit  $W \in \mathbb{R}^{2N \times 2N}$  une matrice  $J$ -symplectique avec  $J \in \mathbb{R}^{2N \times 2N}$  anti-symétrique et inversible. Le spectre de  $W$  est noté  $\Lambda(W)$ .

On considère la décomposition spectrale de  $W$  suivante

$$W = \begin{pmatrix} X_0 & X_1 & X_\infty \end{pmatrix} \begin{pmatrix} W_0 & & \\ & W_1 & \\ & & W_\infty \end{pmatrix} \begin{pmatrix} X_0 & X_1 & X_\infty \end{pmatrix}^{-1} \quad (3.1)$$

avec

- $\Lambda(W_0) = \{\lambda \in \Lambda(W) \text{ telle que } |\lambda| < 1\}$ ,
- $\Lambda(W_1) = \{\lambda \in \Lambda(W) \text{ telle que } |\lambda| = 1\}$ ,
- $\Lambda(W_\infty) = \{\lambda \in \Lambda(W) \text{ telle que } |\lambda| > 1\}$ .

On note  $N_0$  le cardinal de  $\Lambda(W_0)$ , les propriétés de  $J$ -symplecticité de  $W$  impliquent que les cardinaux de  $\Lambda(W_1)$  et  $\Lambda(W_\infty)$  sont respectivement  $2N_1$  avec  $N_1 = N - N_0$  et  $N_0$ .

On pose

$$X = \begin{pmatrix} X_0 & X_1 & X_\infty \end{pmatrix}.$$

Les sous-espaces vectoriels engendrés par les colonnes de  $X$ ,  $X_0$ ,  $X_1$  et  $X_\infty$  sont notés  $\mathcal{X}$ ,  $\mathcal{X}_0$ ,  $\mathcal{X}_1$  et  $\mathcal{X}_\infty$  respectivement et  $\mathbb{P}_0$ ,  $\mathbb{P}_1$  et  $\mathbb{P}_\infty$  désignent les projecteurs spectraux associés respectivement aux sous-espaces invariants  $\mathcal{X}_0$ ,  $\mathcal{X}_1$  et  $\mathcal{X}_\infty$ . Ils sont définis classiquement par

$$\mathbb{P}_0 = X \begin{pmatrix} I & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} X^{-1}, \quad \mathbb{P}_1 = X \begin{pmatrix} 0 & 0 & 0 \\ 0 & I & 0 \\ 0 & 0 & 0 \end{pmatrix} X^{-1}, \quad \mathbb{P}_\infty = X \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & I \end{pmatrix} X^{-1}.$$

Les propriétés de  $J$ -symplecticité de  $W$  impliquent qu'il existe des matrices non singulières  $M$  et  $J_1$  avec  $J_1$  anti-symétrique telles que

$$\begin{pmatrix} X_0 & X_1 & X_\infty \end{pmatrix}^T J \begin{pmatrix} X_0 & X_1 & X_\infty \end{pmatrix} = \begin{pmatrix} & & M \\ & J_1 & \\ -M^T & & \end{pmatrix}.$$

On en déduit que les projecteurs  $\mathbb{P}_0$ ,  $\mathbb{P}_1$  et  $\mathbb{P}_\infty$  vérifient

$$\mathbb{P}_0 = X_0 (X_\infty^T J X_0)^{-1} X_\infty^T J = -X_0 M^{-T} X_\infty^T J,$$

$$\mathbb{P}_1 = X_1 (X_1^T J X_1)^{-1} X_1^T J = X_1 J_1^{-1} X_1^T J,$$

$$\mathbb{P}_\infty = X_\infty (X_0^T J X_\infty)^{-1} X_0^T J = X_\infty M^{-1} X_0^T J.$$

Soient  $\alpha$  et  $\beta$  deux réels strictement positifs tels que  $2\alpha + \beta = 1$  et  $n \in \mathbb{N}^*$ , on suppose que le choix de  $\alpha$  et  $\beta$  permet de définir  $X_0^{(n)}$ ,  $X_1^{(n)}$  et  $X_\infty^{(n)}$  par

$$X_0^{(n)} = W^{-2n} (\beta I + \alpha (W^{-2n} + W^{2n}))^{-1}, \quad (3.2)$$

$$X_1^{(n)} = (\beta I + \alpha (W^{-2n} + W^{2n}))^{-1}, \quad (3.3)$$

$$X_\infty^{(n)} = W^{2n} (\beta I + \alpha (W^{-2n} + W^{2n}))^{-1}, \quad (3.4)$$

autrement dit,

$$X_0^{(n)} = W^{-2^n} X_1^{(n)} \quad \text{et} \quad X_\infty^{(n)} = W^{2^n} X_1^{(n)}. \quad (3.5)$$

On note  $\mathcal{X}_0^{(n)}$ ,  $\mathcal{X}_1^{(n)}$  et  $\mathcal{X}_\infty^{(n)}$  les sous-espaces vectoriels engendrés par les colonnes de  $X_0^{(n)}$ ,  $X_1^{(n)}$  et  $X_\infty^{(n)}$  respectivement.

On définit les matrices  $L_0^{(n)}$ ,  $L_1^{(n)}$  et  $L_\infty^{(n)}$  par

$$\begin{aligned} L_0^{(n)} &= (\beta I + \alpha (W_0^{-2^n} + W_0^{2^n}))^{-1}, \\ L_1^{(n)} &= (\beta I + \alpha (W_1^{-2^n} + W_1^{2^n}))^{-1}, \\ L_\infty^{(n)} &= (\beta I + \alpha (W_\infty^{-2^n} + W_\infty^{2^n}))^{-1}. \end{aligned}$$

En utilisant (3.1), on obtient les décompositions suivantes

$$X_0^{(n)} = X \begin{pmatrix} W_0^{-2^n} L_0^{(n)} & & \\ & W_1^{-2^n} L_1^{(n)} & \\ & & W_\infty^{-2^n} L_\infty^{(n)} \end{pmatrix} X^{-1}, \quad (3.6)$$

$$X_1^{(n)} = X \begin{pmatrix} L_0^{(n)} & & \\ & L_1^{(n)} & \\ & & L_\infty^{(n)} \end{pmatrix} X^{-1}, \quad (3.7)$$

$$X_\infty^{(n)} = X \begin{pmatrix} W_0^{2^n} L_0^{(n)} & & \\ & W_1^{2^n} L_1^{(n)} & \\ & & W_\infty^{2^n} L_\infty^{(n)} \end{pmatrix} X^{-1}. \quad (3.8)$$

**Remarque 4** *Aucune condition de normalisation n'a été imposée sur  $X$ ,  $X_0$ ,  $X_1$  et  $X_\infty$ .*

Tout au long de ce chapitre, le symbole  $\| \cdot \|$  désignera la norme spectrale (ou norme 2).

### 3.2.2 Angles canoniques entre sous-espaces vectoriels

L'analyse de la convergence de l'algorithme de trichotomie présentée dans les sections suivantes, repose sur l'étude de la convergence de suites de sous-espaces vectoriels. Pour réaliser ce travail, nous utilisons la notion d'angles canoniques entre sous-espaces vectoriels de même dimension qui constitue un outil classique permettant d'étudier la convergence d'une suite de sous-espaces vectoriels. En effet, elle permet, entre autres, de définir une distance

sur un ensemble de sous-espaces vectoriels de même dimension.

Soient  $\mathcal{M}_1$  et  $\mathcal{M}_2$  deux sous-espaces vectoriels de  $\mathbb{R}^n$  de même dimension  $m$  et  $M_1, M_2$  deux matrices dont les colonnes forment des bases orthonormées de  $\mathcal{M}_1$  et  $\mathcal{M}_2$  respectivement. Les angles canoniques  $\frac{\pi}{2} \geq \theta_1 \geq \dots \geq \theta_m \geq 0$  entre  $\mathcal{M}_1$  et  $\mathcal{M}_2$  sont définis par

$$\sin(\theta_i) = \sigma_i \quad \text{avec} \quad 1 \leq i \leq m$$

où  $\sigma_m \leq \dots \leq \sigma_1$  sont les  $m$  valeurs singulières non nécessairement nulles de  $(I - \Pi_1)\Pi_2$  avec  $\Pi_1$  et  $\Pi_2$  les projections orthogonales sur  $\mathcal{M}_1$  et  $\mathcal{M}_2$  respectivement.

Dans le cadre de ce travail, l'angle  $\theta_1$  occupe une place prépondérante. En effet, l'application  $d$  définie par

$$d(\mathcal{M}_1, \mathcal{M}_2) = \sin(\theta_1)$$

est une distance sur l'ensemble des sous-espaces vectoriels de  $\mathbb{R}^n$  de dimension  $m$ . De plus, elle vérifie [15]

$$\begin{aligned} d(\mathcal{M}_1, \mathcal{M}_2) &= \|(I - \Pi_1)\Pi_2\| = \|(I - M_1 M_1^T)M_2\| \\ &= \|(I - \Pi_2)\Pi_1\| = \|(I - M_2 M_2^T)M_1\| \\ &= \|\Pi_1 - \Pi_2\| = \|M_1 M_1^T - M_2 M_2^T\| \\ &= \max_{v_1 \in \mathcal{M}_1 - \{0\}} \min_{v_2 \in \mathcal{M}_2} \frac{\|v_1 - v_2\|}{\|v_1\|} \\ &= \max_{v_2 \in \mathcal{M}_2 - \{0\}} \min_{v_1 \in \mathcal{M}_1} \frac{\|v_2 - v_1\|}{\|v_2\|}. \end{aligned}$$

Ces propriétés font de l'application  $d$  un outil efficace et souple pour l'étude de la convergence d'une suite de sous-espaces vectoriels.

L'intérêt de la proposition suivante est purement technique. Elle montre qu'il est possible de choisir des bases orthonormées de  $\mathcal{M}_1$  et  $\mathcal{M}_2$  permettant d'obtenir une décomposition simple de  $\mathcal{M}_1$  en fonction de  $\mathcal{M}_2$  et des angles canoniques.

**Proposition 5** *Si  $d(\mathcal{M}_1, \mathcal{M}_2) < 1$  alors il existe des matrices*

- $\widetilde{M}_2$  dont les colonnes forment une base orthonormée de  $\mathcal{M}_2$ ,
- $S$  symétrique et semi-définie positive,
- $C$  symétrique et définie positive,
- $\widetilde{M}_3$  telle que  $\widetilde{M}_3^T \widetilde{M}_3 = I$

*vérifiant*

$$\begin{aligned} \|S\| &= d(\mathcal{M}_1, \mathcal{M}_2), \\ \lambda_{\min}(C) &= \sqrt{1 - d^2(\mathcal{M}_1, \mathcal{M}_2)}, \\ \text{eng } \widetilde{M}_3 &\perp \mathcal{M}_2, \end{aligned}$$

*telles que*

$$M_1 = \widetilde{M}_2 C + \widetilde{M}_3 S.$$

### Démonstration

La condition  $d(\mathcal{M}_1, \mathcal{M}_2) < 1$  entraîne que  $M_2^T M_1$  inversible. En effet, si  $M_2^T M_1 X = 0$  alors  $(I - M_2 M_2^T) M_1 X = M_1 X$ . On en déduit que  $X = 0$  car  $\|(I - M_2 M_2^T) M_1\| = d(\mathcal{M}_1, \mathcal{M}_2) < 1$ .

Considérons la décomposition en valeurs singulières de  $(I - M_2 M_2^T) M_1$  :

$$(I - M_2 M_2^T) M_1 = U \Sigma V^T$$

avec  $U$  et  $V$  orthogonales et  $\Sigma = \begin{pmatrix} \Sigma_1 \\ 0 \end{pmatrix}$  avec  $\Sigma_1 \in \mathbb{R}^{m \times m}$ . Il est clair que

$$\|\Sigma_1\| = \|(I - M_2 M_2^T) M_1\| = d(\mathcal{M}_1, \mathcal{M}_2).$$

On décompose  $U$  sous la forme

$$U = (U_1, U_2) \text{ avec } \begin{cases} U_1 \in \mathbb{R}^{n \times m}, \\ U_2 \in \mathbb{R}^{n \times (n-m)}. \end{cases}$$

Alors, on a

$$U \Sigma V^T = U_1 \Sigma_1 V^T$$

et donc

$$M_1 = M_2 (M_2^T M_1) + U_1 \Sigma_1 V^T.$$

Puisque  $M_2^T M_1$  est inversible, sa décomposition polaire s'écrit

$$M_2^T M_1 = Q C,$$

où  $Q$  est orthogonale et  $C$  symétrique définie positive.

On a donc

$$M_1 = M_2 Q C + U_1 V^T (V \Sigma_1 V^T)$$

On obtient la décomposition cherchée en posant

$$\begin{aligned} \widetilde{M}_2 &= M_2 Q, \\ \widetilde{M}_3 &= U_1 V^T \end{aligned}$$

et

$$S = V \Sigma_1 V^T.$$

On a ainsi

$$\begin{aligned}\widetilde{M}_2^T \widetilde{M}_2 &= Q^T M_2^T M_2 Q = Q^T Q = I, \\ \widetilde{M}_3^T \widetilde{M}_3 &= V U_1^T U_1 V^T = V V^T = I, \\ \widetilde{M}_3^T \widetilde{M}_2 &= V U_1^T M_2 Q = 0\end{aligned}$$

car  $(I - M_2 M_2^T) U_1 = U_1$  donc  $M_2 M_2^T U_1 = 0$ .

Pour finir, il est clair que  $S$  et  $C$  sont symétriques et semi-définies positives et vérifient

$$\|S\| = \|\Sigma_1\| = d(\mathcal{M}_1, \mathcal{M}_2)$$

et

$$\lambda_{\min}(C) = \sqrt{1 - d^2(\mathcal{M}_1, \mathcal{M}_2)}.$$

En effet, pour tout vecteur  $x$  de norme 1, on a

$$\begin{aligned}\|M_1 x\|^2 &= \|M_2 M_2^T M_1 x + (I - M_2 M_2^T) M_1 x\|^2 \\ &= \|M_2 M_2^T M_1 x\|^2 + \|(I - M_2 M_2^T) M_1 x\|^2\end{aligned}$$

et donc

$$\begin{aligned}1 &= \|M_2^T M_1 x\|^2 + \|(I - M_2 M_2^T) M_1 x\|^2 \\ &= \|Q^T M_2^T M_1 x\|^2 + \|(I - M_2 M_2^T) M_1 x\|^2 \\ &= \|C x\|^2 + \|(I - M_2 M_2^T) M_1 x\|^2.\end{aligned}$$

On en déduit que

$$\begin{aligned}\lambda_{\min}(C)^2 &= \min_{\|x\|=1} \|C x\|^2 \\ &= 1 - \max_{\|x\|=1} \|(I - M_2 M_2^T) M_1 x\|^2 \\ &= 1 - d^2(\mathcal{M}_1, \mathcal{M}_2).\end{aligned}$$

Le fait que  $d(\mathcal{M}_1, \mathcal{M}_2) < 1$  implique donc que  $C$  est définie positive.  $\square$

Il existe beaucoup d'autres résultats concernant les angles canoniques (voir par exemple [13, 22]). Cependant dans ce travail, seules la distance  $d$  et la décomposition proposée dans la proposition 5 sont utilisées.



### 3.2.3 Définition de la convergence (sse)

Afin d'étudier la convergence de la méthode de trichotomie, nous devons définir la convergence d'une suite de sous-espaces vectoriels engendrés par une suite de matrices. En effet, l'application, qui à une matrice, associe son image n'est pas continue ; il n'est donc pas possible de définir la limite des sous-espaces engendrés par une suite de matrices convergente comme étant l'espace engendré par la limite de cette suite de matrices.

Plus précisément, il existe des suites  $(A_n)_{n \geq 0}$  de matrices de  $\mathbb{R}^{p \times p}$  convergeant vers une matrice  $A$  telles que  $\lim_{n \rightarrow \infty} \text{eng } A_n \neq \text{eng } A$ . Pour s'en convaincre, il suffit de considérer la suite  $(A_n)_{n \geq 0}$  définie par  $A_n = \frac{1}{n+1}I$ . Il est clair que  $\text{eng } A_n = \mathbb{R}^p$  pour tout  $n \geq 0$  et donc

$$\lim_{n \rightarrow \infty} \text{eng } A_n = \mathbb{R}^p$$

Or on a

$$\text{eng } \left( \lim_{n \rightarrow \infty} A_n \right) = \text{eng } (0) = \{0\}.$$

D'un autre point de vue, on constate qu'il existe des suites  $(A_n)_{n \geq 0}$  ne convergeant pas alors qu'intuitivement il est facile de donner la limite de  $(\text{eng } A_n)_{n \geq 0}$ . La suite  $(A_n)_{n \geq 0}$  définie par  $A_n = (-1)^n I$  en est un exemple flagrant.

Ce dernier point est certes plus délicat mais il intervient directement dans l'algorithme de trichotomie. L'exemple suivant est simple mais il permet, en partie, de résumer le comportement asymptotique des suites construites par cet algorithme. Soit  $A_n$  définie pour  $n \geq 0$  par

$$A_n = \begin{pmatrix} 1 + \frac{1}{n+1} & & \\ & (-1)^n & \\ & & \frac{1}{n+1} \end{pmatrix}.$$

Les vecteurs de la base canonique sont des vecteurs propres de ces matrices et que pour tout  $n \geq 0$  le rang de ces matrices est 3. Cependant en pratique, il est préférable de s'intéresser au rang numérique d'une matrice défini de la manière suivante :

**Définition 8** *Le rang numérique d'une matrice  $A$  noté  $\text{rang}_{\text{num}}(A)$  est défini par*

$$\text{rang}_{\text{num}}(A) = \min \{ \text{rang}(B) ; \|A - B\| \leq C_A \epsilon_{\text{mach}} \}$$

où

- $\epsilon_{mach}$  est la précision machine,
- $C_A$  est déterminée, classiquement, en fonction de la norme et de la taille de la matrice  $A$ .

On vérifie facilement que, asymptotiquement, le rang numérique des matrices  $A_n$  est 2. De plus, les sous-espaces vectoriels engendrés par les matrices  $B$  de rang 2 vérifiant  $\|A - B\| \leq C_A \epsilon_{mach}$  sont des perturbations des sous-espaces invariants engendrés par les vecteurs

$$e_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} \text{ et } e_2 = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}.$$

Or ces vecteurs engendrent un sous-espace invariant des matrices  $A_n$ . Cette analyse est grossière mais elle montre qu'il est possible, via des suites de matrices divergentes, d'approcher sous certaines conditions des sous-espaces invariants.

Il est clair qu'en général, l'expression des matrices  $A_n$  est nettement plus compliquée et ne permet pas d'utiliser ce type d'analyse trop simple. Afin de contourner ces difficultés, on introduit une notion particulière de limite de suites de sous-espaces vectoriels engendrés par une suite de matrices.

**Définition 9** Soit  $(A_n)_{n \geq 0}$  une suite de matrices de  $\mathbb{R}^{p \times p}$  avec  $p$  un entier non nul et  $\mathcal{A}_n = \text{eng } A_n$ .

La suite  $\mathcal{A}_n$  converge au sens des sous-espaces engendrés vers le sous-espace vectoriel  $\mathcal{A}$  de dimension  $m$  s'il existe un entier  $\tilde{n} > 0$  et deux constantes strictement positives  $K_1$  et  $K_2$  indépendantes de  $n$  vérifiant

$$\|(I - \Pi_n)\Pi\| \xrightarrow{n \rightarrow \infty} 0, \quad (3.9)$$

$$\|(I - \Pi)A_n\| \xrightarrow{n \rightarrow \infty} 0, \quad (3.10)$$

$$\|\Pi A_n\| \leq K_1 \quad \forall n \geq \tilde{n}, \quad (3.11)$$

$$\sigma_m(\Pi A_n) \geq K_2 \quad \forall n \geq \tilde{n} \quad (3.12)$$

où  $\Pi_n$  et  $\Pi$  sont les projections orthogonales sur  $\mathcal{A}_n$  et  $\mathcal{A}$  respectivement et  $\sigma_m(\Pi A_n)$  la  $m$ -ème valeur singulière de  $\Pi A_n$ .

Pour signifier que la convergence est au sens des sous-espaces engendrés ou sse, nous utiliserons la notation

$$\mathcal{A}_n \xrightarrow[n \rightarrow \infty]{(sse)} \mathcal{A}.$$

Le choix de cette définition est motivé par deux conditions : imposer à la suite  $(\|A_n\|)_{n \geq 0}$  d'être bornée et se placer en un point où l'application qui à une matrice associe son image est continue. De plus, comme le montre la proposition suivante, il est consistant avec le cas où la suite  $(A_n)_{n \geq 0}$  converge.

**Proposition 6** *Si la suite  $(A_n)_{n \geq 0}$  converge vers une matrice  $A$  non nulle alors*

$$A_n \xrightarrow[n \rightarrow \infty]{(sse)} A \equiv \text{eng } A.$$

**Démonstration**

On note  $A = MR$  la décomposition QR de rang plein de  $A$ . On a  $\Pi = MM^T$  et

$$\|\Pi A_n\| = \|M^T A_n\| \leq \|R\| + \|A - A_n\| = \|A\| + \|A - A_n\|$$

et

$$\begin{aligned} \sigma_{\min}(\Pi A_n) &= \sigma_{\min}(M^T A_n) \geq \sigma_{\min}(M^T A) - \|M^T(A - A_n)\| \\ &\geq \sigma_{\min}(R) - \|A - A_n\|. \end{aligned}$$

Sachant que  $A_n$  converge vers  $A$ , il existe  $\tilde{n}$  et  $K > 0$  tels que  $\forall n \geq \tilde{n}$ ,

$$\|\Pi A_n\| \leq K \text{ et } \sigma_{\min}(\Pi A_n) \geq \frac{1}{K}. \quad (3.13)$$

De la majoration

$$\|(I - \Pi)A_n\| = \|(I - \Pi)(A_n - A)\| \leq \|A_n - A\|,$$

on déduit la convergence de  $\|(I - \Pi)A_n\|$  vers 0.

Etant donné que  $A_n$  converge vers  $A$ , on a

$$\begin{aligned} \|(I - \Pi_n)\Pi\| &= \max_{v \in \mathcal{A} - \{0\}} \frac{\|(I - \Pi_n)v\|}{\|v\|} = \max_{u ; Au \neq 0} \frac{\|(I - \Pi_n)Au\|}{\|Au\|} \\ &= \max_{u ; Au \neq 0} \min_{v_n \in \mathcal{A}_n} \frac{\|Au - v_n\|}{\|Au\|} \leq \max_{u ; Au \neq 0} \frac{\|Au - A_n u\|}{\|Au\|} \\ &\leq \|A - A_n\| / \sigma_{\min}(R) \end{aligned}$$

impliquant donc que  $\|(I - \Pi_n)\Pi\|$  converge vers 0.  $\square$

La proposition suivante donne des conditions permettant de caractériser la convergence (sse) en terme de suites matricielles.

**Proposition 7** Avec les notations de la définition 9,  $(A_n)_{n \geq 0}$  converge au sens (see) vers  $\mathcal{A}$  de dimension  $m$  si et seulement s'il existe deux suites de matrices  $(\tilde{A}_n)_{n \geq 0}$  de rang  $m$  et  $(\hat{A}_n)_{n \geq 0}$ , deux constantes  $K_3$  et  $K_4$  strictement positives et  $\hat{n}$  vérifiant

$$\tilde{A}_n + \hat{A}_n = A_n \quad \text{et} \quad \tilde{A}_n^T \hat{A}_n = 0, \quad \forall n \geq 0$$

et

$$d(\mathcal{A}, \tilde{\mathcal{A}}_n) \xrightarrow{n \rightarrow \infty} 0, \quad \text{où } \tilde{\mathcal{A}}_n = \text{eng } \tilde{A}_n, \quad (3.14)$$

$$\|A_n\| \leq K_3, \quad \forall n \geq \hat{n}, \quad (3.15)$$

$$\sigma_m(\Pi \tilde{A}_n) \geq K_4, \quad \forall n \geq \hat{n}, \quad (3.16)$$

$$\hat{A}_n \xrightarrow{n \rightarrow \infty} 0. \quad (3.17)$$

### Démonstration

Supposons que  $(A_n)_{n \geq 0}$  converge au sens des sous-espaces engendrés vers  $\mathcal{A}$ .

Soit  $\tilde{\Pi}_n$  le projecteur orthogonal sur  $\tilde{\mathcal{A}}_n \equiv \text{eng}(\Pi_n \Pi)$ . Soit  $\hat{n} \geq \tilde{n}$  tel que  $\forall n \geq \hat{n}$ , (3.9) implique que  $\|(I - \Pi_n)\Pi\| < 1$ . On en déduit  $\tilde{\mathcal{A}}_n$  et  $\mathcal{A}$  sont de même dimension. De plus,  $\tilde{\mathcal{A}}_n$  étant un sous-espace vectoriel de  $\mathcal{A}_n$ , il existe  $\hat{\mathcal{A}}_n$  un sous-espace vectoriel vérifiant

$$\tilde{\mathcal{A}}_n \oplus \hat{\mathcal{A}}_n = \mathcal{A}_n \text{ avec } \tilde{\mathcal{A}}_n \perp \hat{\mathcal{A}}_n.$$

Soit  $\hat{\Pi}_n$  le projecteur orthogonal sur  $\hat{\mathcal{A}}_n$ , il est clair que  $\tilde{\Pi}_n \hat{\Pi}_n = 0$  et que  $\Pi_n$  peut-être décomposé sous la forme

$$\Pi_n = \tilde{\Pi}_n + \hat{\Pi}_n.$$

La définition de  $\tilde{\Pi}_n$  implique que

$$\hat{\Pi}_n \Pi = 0.$$

En effet, si  $v \in \mathcal{A}$  alors  $v = \Pi_n \Pi v + (I - \Pi_n)v$ . Sachant que  $\Pi_n \Pi v \in \tilde{\mathcal{A}}_n$  et  $(I - \Pi_n)v \in \mathcal{A}_n^\perp$ , on montre que  $\hat{\Pi}_n v = 0$  car  $\hat{\Pi}_n \tilde{\mathcal{A}}_n = \{0\}$  et  $\hat{\Pi}_n \mathcal{A}_n^\perp = \{0\}$ .

La relation

$$d(\mathcal{A}, \tilde{\mathcal{A}}_n) = \|(I - \tilde{\Pi}_n)\Pi\| = \|(I - \tilde{\Pi}_n)\Pi - \hat{\Pi}_n \Pi\| = \|(I - \Pi_n)\Pi\|$$

permet de montrer que (3.9) implique (3.14).

De (3.10) et (3.11), on montre qu'il existe  $K_3 > 0$  vérifiant (3.15) car

$$\|A_n\|^2 = \|\Pi A_n\|^2 + \|(I - \Pi)A_n\|^2 \leq K_1 + \|(I - \Pi)A_n\|^2.$$

On pose  $\tilde{A}_n = \tilde{\Pi}_n A_n$  et  $\hat{A}_n = \hat{\Pi}_n A_n$ . De (3.12) et de la relation

$$\Pi \tilde{A}_n = \Pi \tilde{\Pi}_n A_n = \Pi (\Pi_n - \hat{\Pi}_n) A_n = \Pi \Pi_n A_n = \Pi A_n,$$

on déduit (3.16).

Pour vérifier (3.17), on s'appuie sur (3.10), (3.14), (3.15), les propriétés de la distance  $d$  et la majoration

$$\begin{aligned} \|\hat{A}_n\| &= \|\Pi \hat{\Pi}_n \hat{A}_n\| + \|(I - \Pi) \hat{A}_n\| = \|(I - \Pi) A_n\| + \|(I - \Pi) \tilde{\Pi}_n \tilde{A}_n\| \\ &\leq \|(I - \Pi) A_n\| + d(\mathcal{A}, \tilde{\mathcal{A}}_n) \|A_n\|. \end{aligned}$$

#### Réciproque

Dans un premier temps, on considère  $n \geq \hat{n}$ .

Les relations  $A_n = \tilde{A}_n + \hat{A}_n$  et  $\tilde{A}_n^T \hat{A}_n = 0$  impliquent que

$$\text{eng } \tilde{A}_n \oplus \text{eng } \hat{A}_n = \mathcal{A}_n \text{ et } \tilde{\mathcal{A}}_n \perp \hat{\mathcal{A}}_n.$$

On en déduit que  $\tilde{\Pi}_n$  et  $\hat{\Pi}_n$ , les projections orthogonales sur  $\tilde{\mathcal{A}}_n$  et  $\hat{\mathcal{A}}_n$  respectivement, vérifient  $\tilde{\Pi}_n + \hat{\Pi}_n = \Pi_n$  et  $\tilde{\Pi}_n \hat{\Pi}_n = 0$ .

Les matrices  $\tilde{A}_n$  étant de rang  $m$ , on déduit de (3.14) que

$$\tilde{\Pi}_n \xrightarrow{n \rightarrow \infty} \Pi$$

ce qui implique

$$\begin{aligned} \tilde{\Pi}_n \Pi &= \Pi + (\tilde{\Pi}_n - \Pi) \Pi \xrightarrow{n \rightarrow \infty} \Pi \\ \hat{\Pi}_n \Pi &= \hat{\Pi}_n (\Pi - \tilde{\Pi}_n) \xrightarrow{n \rightarrow \infty} 0, \end{aligned}$$

et donc on a

$$\Pi_n \Pi = \tilde{\Pi}_n \Pi + \hat{\Pi}_n \Pi \xrightarrow{n \rightarrow \infty} \Pi.$$

On déduit (3.9) de cette dernière relation car

$$\|(I - \Pi_n) \Pi\| = \|\Pi - \Pi_n \Pi\| \xrightarrow{n \rightarrow \infty} \|\Pi - \Pi\| = 0.$$

On a

$$\begin{aligned} \|(I - \Pi) A_n\| &\leq \|(I - \Pi) \tilde{A}_n\| + \|(I - \Pi) \hat{A}_n\| \\ &\leq \|(I - \Pi) \tilde{\Pi}_n\| \|A_n\| + \|(I - \Pi)\| \|\hat{A}_n\| \end{aligned}$$

et

$$(I - \Pi) \tilde{\Pi}_n = \tilde{\Pi}_n - \left( \tilde{\Pi}_n \Pi \right)^T \xrightarrow{n \rightarrow \infty} \Pi - \Pi^T = 0,$$

donc en se basant sur (3.15) et (3.17), on en déduit (3.10).

Il est clair que (3.15) implique (3.11) avec  $K_1 = K_3$ .

On a

$$\sigma_m(\Pi A_n) \geq \sigma_m(\Pi \tilde{A}_n) - \|\Pi \hat{A}_n\| \geq K_4 - \|\hat{A}_n\|.$$

D'après (3.17), il existe  $\tilde{n} \geq \hat{n}$  tel que  $K_2 = K_4 - \sup_{n \geq \tilde{n}} \|\hat{A}_n\|$  soit strictement positive, ce qui montre (3.12).  $\square$

La proposition suivante montre que la limite au sens (sse) d'une suite est, lorsqu'elle existe, unique.

**Proposition 8** *Soit  $(A_n)_{n \geq 0}$  une suite matricielle telle que*

$$\text{eng } A_n \xrightarrow[n \rightarrow \infty]{(sse)} \mathcal{A}^{(1)} \text{ et } \text{eng } A_n \xrightarrow[n \rightarrow \infty]{(sse)} \mathcal{A}^{(2)}$$

*où  $\mathcal{A}^{(1)}$  et  $\mathcal{A}^{(2)}$  sont deux sous-espaces vectoriels.*

*Alors on a*

$$\mathcal{A}^{(1)} = \mathcal{A}^{(2)}.$$

### Démonstration

D'après la proposition 7, il existe des matrices  $\tilde{A}_n^{(1)}$ ,  $\hat{A}_n^{(1)}$ ,  $\tilde{A}_n^{(2)}$  et  $\hat{A}_n^{(2)}$  vérifiant

$$A_n = \tilde{A}_n^{(1)} + \hat{A}_n^{(1)} = \tilde{A}_n^{(2)} + \hat{A}_n^{(2)}$$

telles que

$$\begin{aligned} d(\mathcal{A}^{(1)}, \tilde{\mathcal{A}}_n^{(1)}) &\xrightarrow{n \rightarrow \infty} 0, & \text{où } \tilde{\mathcal{A}}_n^{(1)} &= \text{eng } \tilde{A}_n^{(1)}, \\ d(\mathcal{A}^{(2)}, \tilde{\mathcal{A}}_n^{(2)}) &\xrightarrow{n \rightarrow \infty} 0, & \text{où } \tilde{\mathcal{A}}_n^{(2)} &= \text{eng } \tilde{A}_n^{(2)} \end{aligned}$$

et  $\hat{A}_n^{(1)}$  et  $\hat{A}_n^{(2)}$  convergent vers 0.

On en déduit que

$$\tilde{A}_n^{(1)} - \tilde{A}_n^{(2)} \xrightarrow{n \rightarrow \infty} 0.$$

En utilisant les propriétés de la distance  $d$ , on a

$$\begin{aligned}
 d(\tilde{\mathcal{A}}_n^{(1)}, \tilde{\mathcal{A}}_n^{(2)}) &= \max_{v_1 \in \tilde{\mathcal{A}}_n^{(1)} - \{0\}} \min_{v_2 \in \tilde{\mathcal{A}}_n^{(2)}} \frac{\|v_1 - v_2\|}{\|v_1\|} \\
 &= \max_{u_1 \in \ker(\tilde{A}_n^{(1)})^\perp - \{0\}} \min_{u_2} \frac{\|\tilde{A}_n^{(1)} u_1 - \tilde{A}_n^{(2)} u_2\|}{\|\tilde{A}_n^{(1)} u_1\|} \\
 &\leq \max_{u_1 \in \ker(\tilde{A}_n^{(1)})^\perp - \{0\}} \frac{\|\tilde{A}_n^{(1)} u_1 - \tilde{A}_n^{(2)} u_1\|}{\|\tilde{A}_n^{(1)} u_1\|} \\
 &\leq \|\tilde{A}_n^{(1)} - \tilde{A}_n^{(2)}\| \max_{u_1 \in \ker(\tilde{A}_n^{(1)})^\perp - \{0\}} \frac{\|u_1\|}{\|\tilde{A}_n^{(1)} u_1\|} \\
 &= \sigma_m(\tilde{A}_n^{(1)})^{-1} \|\tilde{A}_n^{(1)} - \tilde{A}_n^{(2)}\|.
 \end{aligned}$$

La propriété (3.17) implique

$$d(\tilde{\mathcal{A}}_n^{(1)}, \tilde{\mathcal{A}}_n^{(2)}) \xrightarrow{n \rightarrow \infty} 0.$$

On déduit de l'inégalité triangulaire

$$d(\mathcal{A}^{(1)}, \mathcal{A}^{(2)}) \leq d(\mathcal{A}^{(1)}, \tilde{\mathcal{A}}_n^{(1)}) + d(\tilde{\mathcal{A}}_n^{(1)}, \tilde{\mathcal{A}}_n^{(2)}) + d(\tilde{\mathcal{A}}_n^{(2)}, \mathcal{A}^{(2)}),$$

que  $d(\mathcal{A}^{(1)}, \mathcal{A}^{(2)}) = 0$  et donc  $\mathcal{A}^{(1)} = \mathcal{A}^{(2)}$ .  $\square$

La proposition suivante reprend les résultats de la proposition 7 en supprimant une hypothèse concernant les suites  $(\tilde{A}_n)_{n \geq 0}$  et  $(\hat{A}_n)_{n \geq 0}$ .

**Proposition 9** *Avec les notations de la définition 9,  $(A_n)_{n \geq 0}$  converge au sens (see) vers  $\mathcal{A}$  de dimension  $m$  si et seulement s'il existe deux suites de matrices  $(\tilde{A}_n)_{n \geq 0}$  de rang  $m$  et  $(\hat{A}_n)_{n \geq 0}$ , deux constantes  $K_5$  et  $K_6$  strictement positives et  $\check{n}$  vérifiant*

$$\tilde{A}_n + \hat{A}_n = A_n, \quad \forall n \geq 0$$

et

$$d(\mathcal{A}, \tilde{\mathcal{A}}_n) \xrightarrow{n \rightarrow \infty} 0, \quad \text{où } \tilde{\mathcal{A}}_n = \text{eng } \tilde{A}_n, \quad (3.18)$$

$$\|A_n\| \leq K_5, \quad \forall n \geq \check{n}, \quad (3.19)$$

$$\sigma_m(\Pi \tilde{A}_n) \geq K_6, \quad \forall n \geq \check{n}, \quad (3.20)$$

$$\hat{A}_n \xrightarrow{n \rightarrow \infty} 0. \quad (3.21)$$

**Démonstration**

D'après la proposition 7, il suffit de montrer que sous les hypothèses de la proposition 9, les matrices  $\tilde{A}_n$  et  $\hat{A}_n$  permettent de construire deux suites  $(\tilde{B}_n)_{n \geq 0}$  de rang  $m$  et  $(\hat{B}_n)_{n \geq 0}$  vérifiant les hypothèses de la proposition 7. Autrement dit,

$$\tilde{B}_n + \hat{B}_n = A_n \quad \text{et} \quad \tilde{B}_n^T \hat{B}_n = 0, \quad \forall n \geq 0$$

telles qu'il existe une constante  $K_4$  strictement positive et  $\hat{n} \geq \check{n}$  vérifiant

$$\begin{aligned} d(\mathcal{A}, \tilde{\mathcal{B}}_n) &\xrightarrow[n \rightarrow \infty]{} 0, & \text{où } \tilde{\mathcal{B}}_n = \text{eng } \tilde{B}_n, \\ \sigma_m(\Pi \tilde{B}_n) &\geq K_4, & \forall n \geq \hat{n}, \\ \hat{B}_n &\xrightarrow[n \rightarrow \infty]{} 0. \end{aligned}$$

On note  $\tilde{\Pi}_n$  la projection orthogonale sur  $\tilde{\mathcal{A}}_n = \text{eng } \tilde{A}_n$  et on définit les matrices  $\tilde{B}_n$  et  $\hat{B}_n$  par

$$\tilde{B}_n = \tilde{A}_n + \tilde{\Pi}_n \hat{A}_n \quad \text{et} \quad \hat{B}_n = (I - \tilde{\Pi}_n) \hat{A}_n$$

Il est clair que

$$\tilde{B}_n + \hat{B}_n = A_n \quad \text{et} \quad \tilde{B}_n^T \hat{B}_n = 0, \quad \forall n \geq 0.$$

et  $\tilde{\mathcal{B}}_n = \tilde{\mathcal{A}}_n$  ce qui implique

$$d(\mathcal{A}, \tilde{\mathcal{B}}_n) \xrightarrow[n \rightarrow \infty]{} 0.$$

On a

$$\sigma_m(\Pi \tilde{B}_n) \geq \sigma_m(\Pi \tilde{A}_n) - \|\Pi \tilde{\Pi}_n \hat{A}_n\| \geq K_6 - \|\hat{A}_n\|$$

donc il existe  $K_4 > 0$  et  $\check{n}$  tels que  $\forall n \geq \check{n}$ ,

$$\sigma_m(\Pi \tilde{B}_n) \geq K_4.$$

Cette inégalité implique que  $\forall n \geq \check{n}$ ,  $\text{rang } \tilde{B}_n \geq m$ .

De plus,  $\tilde{B}_n = \tilde{\Pi}_n (\tilde{A}_n + \hat{A}_n)$  donc  $\text{rang } \tilde{B}_n \leq m$ . On déduit de ces deux inégalités que  $\forall n \geq \check{n}$ ,

$$\text{rang } \tilde{B}_n = m.$$

□

**Remarque 5** *Il serait intéressant d'étendre la notion de convergence (sse) à un cadre plus général et d'étudier ses principales propriétés. Toutefois la définition et les propriétés données ci-dessus sont suffisantes pour l'étude de l'algorithme de trichotomie.*



### 3.3 Premiers résultats de convergence

L'objet de cette partie est l'étude de la convergence (sse) des suites de sous-espaces vectoriels  $(\mathcal{X}_0^{(n)})_{n \geq 0}$ ,  $(\mathcal{X}_1^{(n)})_{n \geq 0}$  et  $(\mathcal{X}_\infty^{(n)})_{n \geq 0}$ .

L'intérêt de la notion de convergence (sse) apparaît clairement dans cette étude. En effet, les suites  $(\mathcal{X}_0^{(n)})_{n \geq 0}$ ,  $(\mathcal{X}_1^{(n)})_{n \geq 0}$  et  $(\mathcal{X}_\infty^{(n)})_{n \geq 0}$  ne convergent pas au sens classique. Cependant on montre que, sous certaines hypothèses, elles convergent au sens (sse) et ce résultat est suffisant pour justifier l'étape de  $J$ -orthogonalisation expliquée par la suite.

Dans la première partie de cette section, les résultats montrés concernent le comportement des suites  $(X_0^{(n)})_{n \geq 0}$ ,  $(X_1^{(n)})_{n \geq 0}$  et  $(X_\infty^{(n)})_{n \geq 0}$  lorsque  $n$  tend vers l'infini. Il est important de noter qu'ils seront utilisés tout au long de ce travail et notamment dans la deuxième partie de cette section pour montrer que les suites  $(\mathcal{X}_0^{(n)})_{n \geq 0}$ ,  $(\mathcal{X}_1^{(n)})_{n \geq 0}$  et  $(\mathcal{X}_\infty^{(n)})_{n \geq 0}$  convergent au sens (sse).

#### 3.3.1 Etude de la convergence de $(X_0^{(n)})_{n \geq 0}$ , $(X_1^{(n)})_{n \geq 0}$ et $(X_\infty^{(n)})_{n \geq 0}$

Dans cette partie, nous donnons les résultats de convergence concernant les suites  $(X_0^{(n)})_{n \geq 0}$ ,  $(X_1^{(n)})_{n \geq 0}$  et  $(X_\infty^{(n)})_{n \geq 0}$  utilisées par la suite lors de l'étude de convergence des suites de sous-espaces associées.

**Lemme 2** *Pour tout  $p \geq 0$ , on a*

$$\begin{aligned} \|W_0^p\| &\leq \omega_0 \gamma_0^p, & \text{avec } \omega_0 \geq 1 \text{ et } 0 \leq \gamma_0 < 1, \\ \|W_\infty^{-p}\| &\leq \omega_\infty \gamma_\infty^p & \text{avec } \omega_\infty \geq 1 \text{ et } 0 \leq \gamma_\infty < 1. \end{aligned}$$

#### Démonstration

Puisque les valeurs propres de  $W_0$  sont dans le disque unité ouvert, on peut définir la matrice  $\Omega_0$  par

$$\Omega_0 = \sum_{i=0}^{+\infty} (W_0^T)^i W_0^i.$$

D'après [8],  $\Omega_0$  est l'unique solution Hermitienne définie-positive de l'équation de Lyapunov

$$\Omega - W_0^T \Omega W_0 = I_{N_0}.$$

Cette matrice permet d'obtenir la majoration

$$\|W_0^p\| \leq \omega_0 \gamma_0^p \text{ avec } \omega_0 = \sqrt{\|\Omega_0\| \|\Omega_0^{-1}\|} \text{ et } \gamma_0 = \sqrt{1 - \frac{1}{\|\Omega_0\|}}.$$

Pour cela, on considère un vecteur  $x_0$  quelconque et on définit  $x_p$  pour  $p \geq 1$  par

$$x_p = W_0 x_{p-1}.$$

On a

$$\begin{aligned} x_{p+1}^T \Omega_0 x_{p+1} &= x_p^T W_0^T \Omega_0 W_0 x_p = x_p^T (\Omega_0 - I) x_p \\ &= x_p^T \Omega_0 x_p - x_p^T x_p. \end{aligned}$$

Or la matrice  $\Omega_0$  est Hermitienne définie-positive et vérifie pour tout vecteur  $y$

$$y^T \Omega_0 y = y^T y + \sum_{i=1}^{+\infty} (y^T W_0^T)^i W_0^i y \geq y^T y$$

et donc  $y^T \Omega_0 y \geq \|y\|^2$ . Autrement dit, la plus petite valeur propre de  $\Omega_0$  notée  $\lambda_{\min}(\Omega_0)$  est supérieure ou égale à 1. On en déduit, en particulier que  $\|\Omega_0\| \geq 1$  et donc que

$$1 - \frac{1}{\|\Omega_0\|} \geq 0.$$

Or il est clair que

$$x_p^T x_p \geq \frac{1}{\|\Omega_0\|} x_p^T \Omega_0 x_p$$

donc en assemblant ces différents résultats, on peut écrire

$$x_{p+1}^T \Omega_0 x_{p+1} \leq x_p^T \Omega_0 x_p - \frac{1}{\|\Omega_0\|} x_p^T \Omega_0 x_p = \left(1 - \frac{1}{\|\Omega_0\|}\right) x_p^T \Omega_0 x_p.$$

En itérant ce procédé, on obtient la majoration

$$x_p^T \Omega_0 x_p \leq \left(1 - \frac{1}{\|\Omega_0\|}\right)^p x_0^T \Omega_0 x_0$$

d'où on déduit que

$$\lambda_{\min}(\Omega_0) \|x_p\|^2 \leq \|\Omega_0\| \left(1 - \frac{1}{\|\Omega_0\|}\right)^p \|x_0\|^2.$$

ou encore

$$\|x_p\| \leq \omega_0 \gamma_0^p \|x_0\|$$

avec

$$\omega_0 = \sqrt{\frac{\|\Omega_0\|}{\lambda_{\min}(\Omega_0)}} = \sqrt{\|\Omega_0\| \|\Omega_0^{-1}\|} \text{ et } \gamma_0 = \sqrt{1 - \frac{1}{\|\Omega_0\|}}.$$

Or d'après la définition de  $x_p$ , on a

$$\|W_0^p x_0\| \leq \omega_0 \gamma_0^p \|x_0\|$$

ce qui implique que

$$\|W_0^p\| \leq \omega_0 \gamma_0^p$$

car  $x_0$  a été choisi quelconque.

On montre facilement que  $0 \leq \gamma_0 < 1$  en utilisant le fait que  $\|\Omega_0\| \geq 1$  et il est clair que  $\omega_0 \geq 1$  car  $\|\Omega_0\| \|\Omega_0^{-1}\| \geq \|I\| = 1$ .

La démonstration de la deuxième majoration s'obtient de manière tout à fait identique.  $\square$

**Lemme 3** *Il existe deux entiers  $n_0$  et  $n_\infty$  tels que  $\forall n \geq n_0$ ,*

$$\begin{aligned} \|W_0^{2^n} L_0^{(n)}\| &\leq \frac{\omega_0 \gamma_0^{2^{n+1}}}{\alpha - \omega_0 \gamma_0^{2^n}}, \\ \|L_0^{(n)}\| &\leq \frac{\omega_0 \gamma_0^{2^n}}{\alpha - \omega_0 \gamma_0^{2^n}}, \\ \|W_0^{-2^n} L_0^{(n)} - \frac{1}{\alpha} I\| &\leq \frac{1}{\alpha} \frac{\omega_0 \gamma_0^{2^n}}{1 - \omega_0 \gamma_0^{2^n}}, \end{aligned}$$

et  $\forall n \geq n_\infty$

$$\begin{aligned} \|W_\infty^{-2^n} L_\infty^{(n)}\| &\leq \frac{\omega_\infty \gamma_\infty^{2^{n+1}}}{\alpha - \omega_\infty \gamma_\infty^{2^n}}, \\ \|L_\infty^{(n)}\| &\leq \frac{\omega_\infty \gamma_\infty^{2^n}}{\alpha - \omega_\infty \gamma_\infty^{2^n}}, \\ \|W_\infty^{2^n} L_\infty^{(n)} - \frac{1}{\alpha} I\| &\leq \frac{1}{\alpha} \frac{\omega_\infty \gamma_\infty^{2^n}}{\alpha - \omega_\infty \gamma_\infty^{2^n}}. \end{aligned}$$

**Remarque 6** *Ces majorations sont grossières mais elles permettent de fortement simplifier les expressions sans détériorer le caractère de convergence.*

### Démonstration

D'après le lemme 2, il existe  $n_0$  tel que

$$\left\| \frac{\beta}{\alpha} W_0^{2^n} + W_0^{2^{n+1}} \right\| \leq \frac{\beta}{\alpha} \omega_0 \gamma_0^{2^n} + \omega_0 \gamma_0^{2^{n+1}} \leq (1 + \frac{\beta}{\alpha}) \omega_0 \gamma_0^{2^n} \leq \frac{\omega_0}{\alpha} \gamma_0^{2^n} < 1$$

car sachant que  $2\alpha + \beta = 1$  et  $\alpha > 0$ , on montre facilement que  $\frac{\beta}{\alpha} + 1 \leq \frac{1}{\alpha}$ .  
Donc  $\forall n \geq n_0$ , on en déduit la majoration

$$\begin{aligned} \left\| \left( I + \frac{\beta}{\alpha} W_0^{2^n} + W_0^{2^{n+1}} \right)^{-1} \right\| &= \left\| \sum_{i=0}^{\infty} (-1)^i \left( \frac{\beta}{\alpha} W_0^{2^n} + W_0^{2^{n+1}} \right)^i \right\| \\ &\leq \sum_{i=0}^{\infty} \left\| \left( \frac{\beta}{\alpha} W_0^{2^n} + W_0^{2^{n+1}} \right)^i \right\| \\ &\leq \sum_{i=0}^{\infty} \left\| \left( \frac{\beta}{\alpha} W_0^{2^n} + W_0^{2^{n+1}} \right) \right\|^i \\ &\leq \frac{1}{1 - \left\| \left( \frac{\beta}{\alpha} W_0^{2^n} + W_0^{2^{n+1}} \right) \right\|} \\ &\leq \frac{\alpha}{\alpha - \omega_0 \gamma_0^{2^n}}. \end{aligned}$$

Ce qui implique les majorations

$$\begin{aligned} \left\| L_0^{(n)} \right\| &= \left\| \frac{1}{\alpha} W_0^{2^n} \left( I + \frac{\beta}{\alpha} W_0^{2^n} + W_0^{2^{n+1}} \right)^{-1} \right\| \\ &\leq \frac{1}{\alpha} \omega_0 \gamma_0^{2^n} \frac{\alpha}{\alpha - \omega_0 \gamma_0^{2^n}} = \frac{\omega_0 \gamma_0^{2^n}}{\alpha - \omega_0 \gamma_0^{2^n}} \end{aligned}$$

et

$$\begin{aligned} \left\| W_0^{2^n} L_0^{(n)} \right\| &= \left\| \frac{1}{\alpha} W_0^{2^{n+1}} \left( I + \frac{\beta}{\alpha} W_0^{2^n} + W_0^{2^{n+1}} \right)^{-1} \right\| \\ &\leq \frac{1}{\alpha} \omega_0 \gamma_0^{2^{n+1}} \frac{\alpha}{\alpha - \omega_0 \gamma_0^{2^n}} = \frac{\omega_0 \gamma_0^{2^{n+1}}}{\alpha - \omega_0 \gamma_0^{2^n}}. \end{aligned}$$

De même, on montre que  $\forall n \geq n_0$ ,

$$\begin{aligned} \left\| W_0^{-2^n} L_0^{(n)} - \frac{1}{\alpha} I \right\| &\leq \frac{1}{\alpha} \left\| \frac{\beta}{\alpha} W_0^{2^n} + W_0^{2^{n+1}} \right\| \left\| \sum_{k=0}^{\infty} \left( \frac{\beta}{\alpha} W_0^{2^n} + W_0^{2^{n+1}} \right)^k \right\| \\ &\leq \frac{1}{\alpha} \frac{\omega_0 \gamma_0^{2^n}}{\alpha - \omega_0 \gamma_0^{2^n}}. \end{aligned}$$

Les autres majorations s'obtiennent de manière identique.  $\square$

Ce lemme implique les résultats de convergence suivants :

$$\begin{aligned} L_0^{(n)} &\xrightarrow[n \rightarrow \infty]{} 0, & L_{\infty}^{(n)} &\xrightarrow[n \rightarrow \infty]{} 0, \\ W_0^{2^n} L_0^{(n)} &\xrightarrow[n \rightarrow \infty]{} 0, & W_{\infty}^{-2^n} L_{\infty}^{(n)} &\xrightarrow[n \rightarrow \infty]{} 0, \end{aligned}$$

et

$$W_0^{-2n} L_0^{(n)} \xrightarrow{n \rightarrow \infty} \frac{1}{\alpha} I, \quad W_\infty^{2n} L_\infty^{(n)} \xrightarrow{n \rightarrow \infty} \frac{1}{\alpha} I.$$

On en déduit à la vue des expressions (3.6), (3.7) et (3.8) que l'étude de la convergence de  $(X_1^{(n)})_{n \geq 0}$ ,  $(X_0^{(n)})_{n \geq 0}$  et  $(X_\infty^{(n)})_{n \geq 0}$  se réduit finalement à celles de  $(W_1^{-2n} L_1^{(n)})_{n \geq 0}$ ,  $(L_1^{(n)})_{n \geq 0}$  et  $(W_1^{2n} L_1^{(n)})_{n \geq 0}$ . Cependant il est clair que, en général, ces suites ne convergent pas. C'est précisément, pour ce type de difficultés, que la notion de convergence (sse) a été introduite.

Le lemme suivant sera utilisé afin de montrer des résultats de convergence (sse) dans la prochaine sous-section.

**Lemme 4** *Si les hypothèses*

$$W_1 \text{ est diagonalisable sous la forme } W_1 = \Phi_1 D_1 \Phi_1^{-1} \quad (3.22)$$

et

$$\alpha < \frac{1}{4} \quad (3.23)$$

sont vérifiées alors

$$\begin{aligned} \frac{1}{\alpha + (1 - \alpha)\kappa_2(\Phi_1)} &< \sigma_{\min}(W_1^{-2n} L_1^{(n)}) \quad \text{et} \quad \|W_1^{-2n} L_1^{(n)}\| \leq \frac{\kappa_2(\Phi_1)}{1 - 4\alpha}, \\ \frac{1}{1 + 2\alpha(\kappa_2(\Phi_1) - 1)} &< \sigma_{\min}(L_1^{(n)}) \quad \text{et} \quad \|L_1^{(n)}\| \leq \frac{\kappa_2(\Phi_1)}{1 - 4\alpha}, \\ \frac{1}{\alpha + (1 - \alpha)\kappa_2(\Phi_1)} &< \sigma_{\min}(W_1^{2n} L_1^{(n)}) \quad \text{et} \quad \|W_1^{2n} L_1^{(n)}\| \leq \frac{\kappa_2(\Phi_1)}{1 - 4\alpha}, \end{aligned}$$

avec  $\kappa_2(\Phi_1) = \|\Phi_1^{-1}\| \|\Phi_1\|$ .

### Démonstration

Soit  $\mu \in \Lambda(L_1^{(n)})$ , il existe  $\lambda = e^{i\theta} \in \Lambda(W_1)$  avec  $0 \leq \theta < 2\pi$  telle que

$$\mu = (\beta + \alpha(e^{i2n\theta} + e^{-i2n\theta}))^{-1} = (\beta + 2\alpha \cos 2n\theta)^{-1}.$$

On en déduit que si  $\alpha < 1/4$  alors

$$|\mu| \leq \frac{1}{\beta - 2\alpha} \leq \frac{1}{1 - 4\alpha}.$$

En utilisant (3.22), on a

$$\|L_1^{(n)}\| \leq \|\Phi_1\| \|\Phi_1^{-1}\| \max_{\mu \in \Lambda(L_1^{(n)})} |\mu| \leq \kappa_2(\Phi_1) \frac{1}{1 - 4\alpha}$$

et

$$\begin{aligned}\sigma_{min}^{-1} \left( L_1^{(n)} \right) &= \beta + \alpha \left\| W_1^{2^n} + W_1^{-2^n} \right\| \\ &\leq \beta + \alpha \kappa_2(\Phi_1) \max_{\lambda \in \Lambda(W_1)} \left| \lambda^{2^n} + \lambda^{-2^n} \right| \leq \beta + 2\alpha \kappa_2(\Phi_1).\end{aligned}$$

Sachant que  $\kappa_2(\Phi_1) \geq 1$ , on a  $\beta + 2\alpha \kappa_2(\Phi_1) > 1$ .

Soit  $\mu \in \Lambda \left( W_1^{-2^n} L_1^{(n)} \right)$ , il existe  $\lambda \in \Lambda(W_1)$  telle que  $\mu = \frac{1}{\beta \lambda^{2^n} + \alpha(1 + \lambda^{4^n})}$ .

On en déduit que  $|\mu| \leq \frac{1}{\beta - 2\alpha} = \frac{1}{1 - 4\alpha}$  et par la suite

$$\left\| W_1^{-2^n} L_1^{(n)} \right\| \leq \frac{\kappa_2(\Phi_1)}{1 - 4\alpha}$$

et

$$\sigma_{min}^{-1} \left( W_1^{-2^n} L_1^{(n)} \right) \leq \alpha + \kappa_2(\Phi_1) (\alpha + \beta) = \alpha + \kappa_2(\Phi_1) (1 - \alpha).$$

Or  $\alpha < 1/4$  donc  $\alpha + \kappa_2(\Phi_1) (1 - \alpha) > 0$ .

Les deux autres inégalités s'obtiennent de manière identique.  $\square$

### 3.3.2 Convergence au sens (sse)

Le théorème suivant ne reprend qu'une partie des résultats de cette section. En particulier, il ne donne aucune indication quant à la vitesse de convergence des différentes suites. Son principal intérêt est de montrer la nécessité de l'étape de  $J$ -orthogonalisation qui est l'objet de l'étude menée dans la section suivante.

**Théorème 10** *Sous les hypothèses (3.22) et (3.23), on a*

$$\begin{aligned}\mathcal{X}_0^{(n)} &\xrightarrow[n \rightarrow \infty]{(sse)} \mathcal{X}_0 + \mathcal{X}_1, \\ \mathcal{X}_1^{(n)} &\xrightarrow[n \rightarrow \infty]{(sse)} \mathcal{X}_1, \\ \mathcal{X}_\infty^{(n)} &\xrightarrow[n \rightarrow \infty]{(sse)} \mathcal{X}_\infty + \mathcal{X}_1.\end{aligned}$$

#### Démonstration

Seule la démonstration du premier résultat est donnée, les deux autres étant

identiques.

Elle repose sur la proposition 9 via les lemmes 3 et 4. On décompose  $X_0^{(n)}$  sous la forme  $X_0^{(n)} = \tilde{A}_n + \hat{A}_n$  avec

$$\tilde{A}_n = (X_0 \ X_1) \begin{pmatrix} W_0^{-2^n} L_0^{(n)} & 0 & 0 \\ 0 & W_1^{-2^n} L_1^{(n)} & 0 \end{pmatrix} X^{-1}$$

et

$$\hat{A}_n = X_\infty \begin{pmatrix} 0 & 0 & W_\infty^{-2^n} L_\infty^{(n)} \end{pmatrix} X^{-1}.$$

il suffit de montrer que, sous les hypothèses (3.22) et (3.23), il existe  $n \geq n^* \geq \max\{n_0, n_\infty\}$  tel que pour  $n \geq n^*$ , les suites  $\tilde{A}_n$  et  $\hat{A}_n$  vérifient les conditions de la proposition 9 impliquant donc que

$$\mathcal{X}_0^{(n)} \xrightarrow[n \rightarrow \infty]{(sse)} \mathcal{X}_0 + \mathcal{X}_1.$$

Il est clair qu'il existe  $n^*$  tel que pour  $n \geq n^*$ , les matrices  $W_0^{-2^n} L_0^{(n)}$  et  $W_1^{-2^n} L_1^{(n)}$  sont de rang plein. En effet, d'après le lemme 4,  $W_1^{-2^n} L_1^{(n)}$  est de rang plein et d'après le lemme 3,  $W_0^{-2^n} L_0^{(n)}$  vérifie pour  $n \geq n^*$

$$\begin{aligned} \sigma_{\min} \left( W_0^{-2^n} L_0^{(n)} \right) &\geq \sigma_{\min} \left( \frac{1}{\alpha} I \right) - \frac{1}{\alpha} \frac{\omega_0 \gamma_0^{2^n}}{\alpha - \omega_0 \gamma_0^{2^n}} \\ &= \frac{1}{\alpha} \left( 1 - \frac{\omega_0 \gamma_0^{2^n}}{\alpha - \omega_0 \gamma_0^{2^n}} \right) > 0. \end{aligned}$$

On en déduit que pour  $n \geq n^*$ , on a

$$\text{eng } \tilde{A}_n = \tilde{\mathcal{A}}_n = \mathcal{X}_0 + \mathcal{X}_1.$$

Il est donc évident que

$$d \left( \mathcal{X}_0 + \mathcal{X}_1, \tilde{\mathcal{A}}_n \right) \xrightarrow[n \rightarrow \infty]{} 0.$$

De plus, on a

$$\Pi \tilde{A}_n = \tilde{A}_n$$

où  $\Pi$  est la projection orthogonale sur  $\mathcal{X}_0 + \mathcal{X}_1$ . Sachant que  $\mathcal{X}_0 + \mathcal{X}_1$  est de dimension  $N_0 + 2N_1$ , on obtient que pour  $n \geq n^*$ , il existe une constante  $K_6 > 0$  telle que

$$\sigma_{N_0+2N_1} \left( \Pi \tilde{A}_n \right) \geq K_6$$

car on montre que

$$\begin{aligned}\sigma_{N_0+2N_1}(\Pi\tilde{A}_n) &= \sigma_{N_0+2N_1}(\tilde{A}_n) \\ &\geq \frac{1}{\kappa_2(X)} \min \left\{ \sigma_{\min}(W_0^{-2^n} L_0^{(n)}), \sigma_{\min}(W_1^{-2^n} L_1^{(n)}) \right\} \\ &\geq \frac{1}{\kappa_2(X)} \min \left\{ \frac{1}{\alpha} \left( 1 - \frac{\omega_0 \gamma_0^{2^n}}{\alpha - \omega_0 \gamma_0^{2^n}} \right), \frac{1}{1 + 2\alpha(\kappa_2(\Phi_1) - 1)} \right\}\end{aligned}$$

en utilisant le fait que

$$\begin{aligned}\sigma_{\min}(X_0 \ X_1) &= \sigma_{\min} \left( X \begin{pmatrix} I & 0 \\ 0 & I \end{pmatrix} \right) \\ &= \min_{\|v\|=1} \left\| X \begin{pmatrix} I & 0 \\ 0 & I \end{pmatrix} v \right\| \\ &\geq \min_{\|v\|=1} \|Xv\| = \sigma_{\min}(X)\end{aligned}$$

et donc

$$\begin{aligned}\sigma_{\min}(X_0 \ X_1) \sigma_{\min}(X^{-1}) &\geq \sigma_{\min}(X) \sigma_{\min}(X^{-1}) \\ &= \frac{1}{\|X^{-1}\|} \frac{1}{\|X\|} = \frac{1}{\kappa_2(X)}.\end{aligned}$$

Il est évident d'après les lemmes 3 et 4 qu'il existe  $K_5 > 0$  telle que

$$\|A_n\| \leq K_5$$

car

$$\begin{aligned}\|A_n\| &\leq \|X\| \max \left\{ \|W_0^{-2^n} L_0^{(n)}\|, \|W_1^{-2^n} L_1^{(n)}\|, \|W_\infty^{-2^n} L_\infty^{(n)}\| \right\} \|X^{-1}\| \\ &\leq \kappa(X) \max \left\{ \frac{1}{\alpha} + \frac{\omega_0 \gamma_0^{2^n}}{\alpha - \omega_0 \gamma_0^{2^n}}, \frac{\kappa_2(\phi_1)}{1 - 4\alpha}, \frac{\omega_\infty \gamma_\infty^{2^{n+1}}}{\alpha - \omega_\infty \gamma_\infty^{2^n}} \right\}.\end{aligned}$$

La dernière condition de la proposition 9 est directement obtenue en remarquant que

$$\|\hat{A}_n\| \leq \kappa(X) \|L_\infty^{(n)}\| \leq \kappa(X) \frac{\omega_\infty \gamma_\infty^{2^{n+1}}}{\alpha - \omega_\infty \gamma_\infty^{2^n}}$$



avec  $\gamma_\infty < 1$  et  $\omega_\infty \geq 1$ .  $\square$

Ce théorème montre clairement le comportement des suites de sous-espaces vectoriels construites par l'algorithme de trichotomie spectrale. On constate, en particulier, que les suites  $(\mathcal{X}_0^{(n)})_{n \geq 0}$  et  $(\mathcal{X}_\infty^{(n)})_{n \geq 0}$  ne convergent pas uniquement vers  $\mathcal{X}_0$  et  $\mathcal{X}_\infty$ . Afin d'en déduire des approximations de  $\mathcal{X}_0$  et  $\mathcal{X}_\infty$ , il est nécessaire de recourir à une étape de post-traitement afin d'éliminer les parties correspondantes au sous-espace  $\mathcal{X}_1$  dans  $X_0^{(n)}$  et  $X_\infty^{(n)}$ . Une manière commode de réaliser cette élimination est basée sur l'utilisation du projecteur spectral  $\mathbb{P}_1$  associé à  $\mathcal{X}_1$ . En effet,  $\mathbb{P}_1$  vérifie

$$(I - \mathbb{P}_1) \mathcal{X}_0 = \mathcal{X}_0, \quad (I - \mathbb{P}_1) \mathcal{X}_1 = 0, \quad (I - \mathbb{P}_1) \mathcal{X}_\infty = \mathcal{X}_\infty.$$

Ainsi en multipliant  $X_0^{(n)}$  et  $X_\infty^{(n)}$  par  $(I - \mathbb{P}_1)$ , il est possible d'éliminer les composantes associées à  $\mathcal{X}_1$  sans modifier celles associées à  $\mathcal{X}_0$  et  $\mathcal{X}_\infty$ . Cette étape sera désignée sous le nom de "étape de J-orthogonalisation".

Cependant cette étape nécessite de connaître  $\mathbb{P}_1$ . Autrement dit, elle implique que  $\mathcal{X}_1$  doit être calculé exactement, ce qui n'est évidemment pas le cas. On ne connaît effectivement qu'une approximation de  $\mathcal{X}_1$  et donc de  $\mathbb{P}_1$ .

La section suivante est consacrée entièrement à l'étude de la précision de l'estimation de  $\mathcal{X}_1$  car elle se révèle être la clef de voûte de l'algorithme de trichotomie spectrale. En effet, une trop grande imprécision de cette estimation se répercuterait inmanquablement sur les approximations de  $\mathcal{X}_0$  et  $\mathcal{X}_\infty$  obtenues via une étape de J-orthogonalisation fort imprécise.

### 3.4 Approximations de $\mathcal{X}_1$ et $\mathbb{P}_1$

Cette section est consacrée à l'étude de la qualité d'une approximation  $\tilde{X}_1^{(n)}$  de  $X_1$  obtenue via  $X_1^{(n)}$  ainsi qu'à celle du projecteur  $\tilde{\mathbb{P}}_1^{(n)}$  défini par

$$\tilde{\mathbb{P}}_1^{(n)} = \tilde{X}_1^{(n)} \left( (\tilde{X}_1^{(n)})^T J \tilde{X}_1^{(n)} \right)^{-1} (\tilde{X}_1^{(n)})^T J.$$

Cette étape détermine en grande partie la qualité de l'algorithme de trichotomie spectrale. En effet, le calcul des approximations de  $\mathbb{P}_0$ ,  $\mathbb{P}_1$  et  $\mathbb{P}_\infty$  étant réalisé en utilisant directement  $\tilde{X}_1^{(n)}$ , il est nécessaire, autant que faire se peut, d'assurer que  $\tilde{X}_1^{(n)}$  est une bonne approximation de  $X_1^{(n)}$ .

Différentes manières de calculer  $\tilde{X}_1^{(n)}$  à partir de  $X_0^{(n)}$ ,  $X_1^{(n)}$  et  $X_\infty^{(n)}$  sont envisageables. En se basant sur les résultats du théorème 10, on constate que  $\mathcal{X}_1$  peut être approchée en se basant sur l'intersection de  $\mathcal{X}_0^{(n)}$  et  $\mathcal{X}_\infty^{(n)}$ . Il est

également possible de déterminer  $\tilde{X}_1^{(n)}$  via un calcul d'éléments spectraux de  $X_1^{(n)}$  car  $\mathcal{X}_1$  est un sous-espace invariant de  $X_1^{(n)}$ . Nous avons choisi d'utiliser une autre voie basée sur une décomposition en valeurs singulières de  $X_1^{(n)}$  car cette décomposition est très stable numériquement.

La sous-section suivante est consacrée au calcul de  $\tilde{X}_1^{(n)}$  et à l'estimation de la qualité de cette approximation. Des résultats concernant l'estimation de l'erreur obtenue en remplaçant  $\mathbb{P}_1$  par  $\tilde{\mathbb{P}}_1^{(n)}$  terminent cette section.

### 3.4.1 Approximation de $\mathcal{X}_1$

La détermination de  $\tilde{X}_1^{(n)}$  est réalisée en utilisant une décomposition en valeurs singulières de  $X_1^{(n)}$  sous la forme :

$$X_1^{(n)} = U^{(n)} \Sigma^{(n)} (V^{(n)})^T \quad \text{avec} \quad \begin{cases} U^{(n)} \text{ et } V^{(n)} \text{ orthogonales,} \\ \Sigma^{(n)} = \text{diag} \left( \sigma_1^{(n)}, \dots, \sigma_{2N}^{(n)} \right) \\ \text{avec } \sigma_1^{(n)} \geq \dots \geq \sigma_{2N}^{(n)} \geq 0. \end{cases}$$

La première étape ne consiste pas à déterminer  $\tilde{X}_1^{(n)}$  mais uniquement la dimension de  $X_1$  qui est inconnue a priori. La proposition suivante montre que si  $n$  est suffisamment grand alors seules  $2N_1$  valeurs singulières de  $X_1^{(n)}$  ne convergent pas vers 0.

**Proposition 10** Soit  $l^{(n)}$  définie par :

$$l^{(n)} = \frac{1}{\kappa_2(X)} \frac{1}{1 + 2\alpha(\kappa_2(\Phi_1) - 1)} - \kappa_2(X) \frac{\omega \gamma^{2^n}}{\alpha - \omega \gamma^{2^n}}$$

où  $\omega = \max \{\omega_0, \omega_\infty\}$  et  $0 \leq \gamma = \max \{\gamma_0, \gamma_\infty\} < 1$ .

Il existe  $l > 0$  et un entier  $n_1$  tel que pour tout  $n \geq n_1$ ,

$$\sigma_{2N_1}^{(n)} - \sigma_{2N_1+1}^{(n)} \geq l^{(n)} > l > 0.$$

#### Démonstration

On découpe  $U^{(n)}$ ,  $V^{(n)}$  et  $\Sigma^{(n)}$  sous la forme

$$U^{(n)} = \begin{pmatrix} U_1^{(n)} & U_2^{(n)} \end{pmatrix}, \quad V^{(n)} = \begin{pmatrix} V_1^{(n)} & V_2^{(n)} \end{pmatrix} \quad \text{et} \quad \Sigma^{(n)} = \begin{pmatrix} \Sigma_1^{(n)} & \\ & \Sigma_2^{(n)} \end{pmatrix}$$

avec  $U_1^{(n)}, V_1^{(n)} \in \mathbb{R}^{2N \times 2N_1}$  et  $\Sigma_1^{(n)} \in \mathbb{R}^{2N_1 \times 2N_1}$ .

D'après le théorème de Schmidt-Mirsky [26, p.208],  $U_1^{(n)} \Sigma_1^{(n)} (V_1^{(n)})^T$  vérifie

$$\left\| X_1^{(n)} - U_1^{(n)} \Sigma_1^{(n)} (V_1^{(n)})^T \right\| = \min_{\text{rang}(A)=2N_1} \left\| X_1^{(n)} - A \right\| = \sigma_{2N_1+1}.$$

En particulier,  $L_1^{(n)}$  étant de rang  $2N_1$ , on a pour  $n \geq \max\{n_0, n_\infty\}$ ,

$$\begin{aligned}\sigma_{2N_1+1} &\leq \left\| X_1^{(n)} - X \begin{pmatrix} 0 & & \\ & L_1^{(n)} & \\ & & 0 \end{pmatrix} X^{-1} \right\| \\ &\leq \kappa_2(X) \max \left\{ \|L_0^{(n)}\|, \|L_\infty^{(n)}\| \right\} \\ &\leq \kappa_2(X) \frac{\omega\gamma^{2^n}}{\alpha - \omega\gamma^{2^n}}\end{aligned}$$

d'après le lemme 3.

D'autre part, d'après le théorème de Courant-Fischer [13, p. 411 et 428], on a

$$\sigma_{2N_1} = \max_{\dim \mathcal{A}=2N_1} \min_{u \neq 0, u \in \mathcal{A}} \frac{\|X_1^{(n)}u\|}{\|u\|}.$$

Ce théorème implique donc que

$$\begin{aligned}\sigma_{2N_1} &\geq \min_{u \neq 0, u \in \mathcal{X}_1} \frac{\|X_1^{(n)}u\|}{\|u\|} = \min_{v \neq 0} \frac{\|X_1 L_1^{(n)}v\|}{\|X_1 v\|} \\ &\geq \frac{\sigma_{\min}(L_1^{(n)})}{\kappa_2(X_1)} \geq \frac{\sigma_{\min}(L_1^{(n)})}{\kappa_2(X)}\end{aligned}$$

car  $\sigma_{\min}(X_1) \geq \sigma_{\min}(X)$  et  $\|X_1\| \leq \|X\|$ . Or d'après le lemme 4,

$$\sigma_{\min}(L_1^{(n)}) \geq (1 + 2\alpha(\kappa_2(\Phi_1) - 1))^{-1}$$

donc

$$\sigma_{2N_1} \geq \frac{1}{\kappa_2(X)} \frac{1}{1 + 2\alpha(\kappa_2(\Phi_1) - 1)}.$$

On en déduit donc

$$\sigma_{2N_1}^{(n)} - \sigma_{2N_1+1}^{(n)} \geq \frac{1}{\kappa_2(X)} \frac{1}{1 + 2\alpha(\kappa_2(\Phi_1) - 1)} - \kappa_2(X) \frac{\omega\gamma^{2^n}}{\alpha - \omega\gamma^{2^n}} = l^{(n)}$$

La suite  $l^{(n)}$  est positive croissante à partir d'un entier  $n_1$  et majorée par  $\frac{1}{\kappa_2(X)} \frac{1}{1 + 2\alpha(\kappa_2(\Phi_1) - 1)}$ . Alors  $\forall n \geq n_1$ ,  $l^{(n)} > l = l^{(n_1)}$ .  $\square$

Dans la suite, on suppose que la séparation entre les  $2N_1$  plus grandes valeurs singulières de  $X_1^{(n)}$  et les autres permet de déterminer la valeur de

$N_1$ . On décompose alors les matrices  $U^{(n)}$ ,  $V^{(n)}$  et  $\Sigma^{(n)}$  sous la forme :

$$U^{(n)} = \begin{pmatrix} U_1^{(n)} & U_2^{(n)} \end{pmatrix}, \quad V^{(n)} = \begin{pmatrix} V_1^{(n)} & V_2^{(n)} \end{pmatrix}, \quad \Sigma^{(n)} = \begin{pmatrix} \Sigma_1^{(n)} & \\ & \Sigma_2^{(n)} \end{pmatrix}$$

où  $U_1^{(n)}$  et  $V_1^{(n)}$  sont des matrices de  $\mathbb{R}^{2N \times 2N_1}$ .

On définit  $\tilde{X}_1^{(n)}$  l'approximation de  $X_1$  obtenue à l'itération  $n$  par

$$\tilde{X}_1^{(n)} = U_1^{(n)}$$

et on note  $\tilde{\mathcal{X}}_1^{(n)}$  le sous-espace vectoriel engendré par les colonnes de  $\tilde{X}_1^{(n)}$ .

Le théorème suivant montre que, sous certaines conditions,  $\tilde{\mathcal{X}}_1^{(n)}$  tend vers  $\mathcal{X}_1$  de manière hypergéométrique.

**Théorème 11** *Sous les hypothèses (3.22) et (3.23), si  $n \geq n_1$  alors*

$$d(\mathcal{X}_1, \tilde{\mathcal{X}}_1^{(n)}) \leq K_X \frac{\omega \gamma^{2^n}}{\alpha - \omega \gamma^{2^n}}$$

avec  $K_X$  une constante indépendante de  $n$  donnée par

$$K_X = \frac{\kappa_2^2(X)}{l^2} \left( \frac{2\kappa_2(\Phi_1)}{1 - 4\alpha} + 1 \right).$$

La démonstration de ce théorème repose sur le théorème 3.6 de [26, p. 251] que nous rappelons ici, en théorème 12 pour plus de commodité.

**Théorème 12** *Soit  $A \in \mathbb{R}^{n \times n}$  une matrice symétrique admettant la bloc-diagonalisation*

$$A = \begin{pmatrix} U_1 & U_2 \end{pmatrix} \begin{pmatrix} A_1 & \\ & A_2 \end{pmatrix} \begin{pmatrix} U_1 & U_2 \end{pmatrix}^T$$

avec  $U_1 \in \mathbb{R}^{n \times k}$  et  $U_2 \in \mathbb{R}^{n \times n-k}$  telle que  $\begin{pmatrix} U_1 & U_2 \end{pmatrix}$  est orthogonale.

Soit  $Z \in \mathbb{R}^{n \times k}$  une matrice dont les colonnes sont orthonormées et  $M \in \mathbb{R}^{k \times k}$  symétrique, on définit  $R$  par

$$R = AZ - ZM.$$

S'il existe  $\delta > 0$  et  $a < b$  vérifiant

$$\Lambda(M) \subset [a, b]$$

et

$$\Lambda(A_2) \subset ]-\infty, a - \delta] \cup [b + \delta, +\infty[$$

alors

$$d(\mathcal{Z}, \mathcal{U}_1) \leq \frac{\|R\|}{\delta}$$

avec  $\mathcal{Z} = \text{Im} Z$  et  $\mathcal{U}_1 = \text{Im} U_1$ .

### Démonstration du théorème 11

Les factorisations QR de rang plein des matrices  $X_1$  et  $(X_0 \ X_\infty)$ , on obtient des matrices  $Y_1 \in \mathbb{R}^{2N \times 2N_1}$  et  $Y_2 \in \mathbb{R}^{2N \times 2(N-N_1)}$  vérifiant  $Y_1^T Y_1 = I$  et  $Y_2^T Y_2 = I$  telles que

$$(X_1 \ (X_0 \ X_\infty)) = (Y_1 \ Y_2) \begin{pmatrix} R_1 & \\ & R_2 \end{pmatrix} \quad (3.24)$$

où  $R_1 \in \mathbb{R}^{2N_1 \times 2N_1}$  et  $R_2 \in \mathbb{R}^{2(N-N_1) \times 2(N-N_1)}$  sont inversibles.

De même, en utilisant des factorisations QR de rang plein, on détermine des matrices  $Z_1 \in \mathbb{R}^{2N \times 2N_1}$  et  $Z_2 \in \mathbb{R}^{2N \times 2(N-N_1)}$  vérifiant  $Z_1^T Z_1 = I$  et  $Z_2^T Z_2 = I$  telles que

$$(X_1 \ (X_0 \ X_\infty))^{-T} = (Z_1 \ Z_2) \begin{pmatrix} S_1 & \\ & S_2 \end{pmatrix} \quad (3.25)$$

où  $S_1 \in \mathbb{R}^{2N_1 \times 2N_1}$  et  $S_2 \in \mathbb{R}^{2(N-N_1) \times 2(N-N_1)}$  sont inversibles.

On montre aisément que

$$\begin{aligned} Z_2^T Y_1 &= 0, \\ Z_1^T Y_2 &= 0, \\ S_1^T Z_1^T Y_1 R_1 &= I, \\ S_2^T Z_2^T Y_2 R_2 &= I. \end{aligned}$$

On a

$$\begin{aligned} X_1^{(n)} &= (Y_1 \ Y_2) \begin{pmatrix} R_1 & \\ & R_2 \end{pmatrix} \begin{pmatrix} L_1^{(n)} & \\ & \begin{pmatrix} L_0^{(n)} & \\ & L_\infty^{(n)} \end{pmatrix} \end{pmatrix} \begin{pmatrix} S_1 & \\ & S_2 \end{pmatrix}^T (Z_1 \ Z_2)^T \\ &= (Y_1 \ Y_2) \begin{pmatrix} K_1 & \\ & K_2 \end{pmatrix} (Z_1 \ Z_2)^T \end{aligned}$$

avec  $K_1 = R_1 L_1^{(n)} S_1^T$  et  $K_2 = R_2 \begin{pmatrix} L_0^{(n)} & \\ & L_\infty^{(n)} \end{pmatrix} S_2^T$ .

Soient  $\Theta$  et  $\Gamma$  les matrices définies par

$$\begin{pmatrix} Y_1 & Y_2 \end{pmatrix}^T \begin{pmatrix} Y_1 & Y_2 \end{pmatrix} = \begin{pmatrix} I & \Theta \\ \Theta^T & I \end{pmatrix}$$

et

$$\begin{pmatrix} Z_1 & Z_2 \end{pmatrix}^T \begin{pmatrix} Z_1 & Z_2 \end{pmatrix} = \begin{pmatrix} I & \Gamma \\ \Gamma^T & I \end{pmatrix}.$$

On vérifie aisément que  $\|\Theta\| \leq 1$  et  $\|\Gamma\| \leq 1$ .

On a donc

$$\begin{aligned} X_1^{(n)} (X_1^{(n)})^T &= \begin{pmatrix} Y_1 & Y_2 \end{pmatrix} \begin{pmatrix} K_1 & \\ & K_2 \end{pmatrix} \begin{pmatrix} I & \Gamma \\ \Gamma^T & I \end{pmatrix} \begin{pmatrix} K_1 & \\ & K_2 \end{pmatrix}^T \begin{pmatrix} Y_1 & Y_2 \end{pmatrix}^T \\ &= \begin{pmatrix} Y_1 & Y_2 \end{pmatrix} \begin{pmatrix} K_1 K_1^T & K_1 \Gamma K_2^T \\ K_2 \Gamma^T K_1^T & K_2 K_2^T \end{pmatrix} \begin{pmatrix} Y_1 & Y_2 \end{pmatrix}^T. \end{aligned}$$

Ce qui implique que

$$X_1^{(n)} (X_1^{(n)})^T Y_1 = \begin{pmatrix} Y_1 & Y_2 \end{pmatrix} \begin{pmatrix} K_1 K_1^T + K_1 \Gamma K_2^T \Theta^T \\ K_2 \Gamma^T K_1^T + K_2 K_2^T \Theta^T \end{pmatrix} \quad (3.26)$$

On note  $R$  le résidu matriciel défini par

$$R = X_1^{(n)} (X_1^{(n)})^T Y_1 - Y_1 K_1 K_1^T. \quad (3.27)$$

On déduit de (3.26) que  $R$  vérifie

$$R = Y_1 K_1 \Gamma K_2 \Theta^T + Y_2 (K_2 \Gamma^T K_1 + K_2 K_2^T \Theta^T).$$

Par ailleurs, on se donne une bloc-diagonalisation de  $X_1^{(n)} (X_1^{(n)})^T$  issue d'une décomposition en valeurs singulières

$$X_1^{(n)} (X_1^{(n)})^T = \begin{pmatrix} U_1^{(n)} & U_2^{(n)} \end{pmatrix} \begin{pmatrix} \left( \Sigma_1^{(n)} \right)^1 & \\ & \left( \Sigma_2^{(n)} \right)^2 \end{pmatrix} \begin{pmatrix} U_1^{(n)} & U_2^{(n)} \end{pmatrix}^T.$$

Pour appliquer le théorème 12, il faut minorer  $\delta$ , qui dans ce contexte, vérifie

$$\delta \geq \sigma_{\min}^2(K_1) - \left\| \left( \Sigma_2^{(n)} \right)^2 \right\| = \sigma_{\min}^2(K_1) - \left( \sigma_{2N_1+1}^{(n)} \right)^2.$$

Or, on a

$$\sigma_{\min}(K_1) = \sigma_{\min}(R_1 L_1^{(n)} S_1^T) \geq \sigma_{\min}(R_1) \sigma_{\min}(L_1^{(n)}) \sigma_{\min}(S_1).$$

Il est facile de voir, d'après (3.24) et (3.25), que

$$\sigma_{\min}(R_1) \sigma_{\min}(S_1) \geq \frac{1}{\kappa_2(X)}.$$

Donc d'après le lemme 4,

$$\sigma_{\min}(K_1) \geq \frac{1}{\kappa_2(X)} \frac{1}{1 + 2\alpha(\kappa_2(\Phi_1) - 1)}.$$

En répétant la démonstration de la proposition 10, on montre que

$$\sigma_{2N_1+1}^{(n)} \leq \kappa_2(X) \frac{\omega \gamma^{2^n}}{\alpha - \omega \gamma^{2^n}}.$$

Par conséquent,

$$\begin{aligned} \delta &\geq \frac{1}{\kappa_2^2(X)} \frac{1}{1 + 2\alpha(\kappa_2(\Phi_1) - 1)^2} - \kappa_2^2(X) \left( \frac{\omega \gamma^{2^n}}{\alpha - \omega \gamma^{2^n}} \right)^2 \\ &\geq (l^n)^2 \\ &\geq l^2. \end{aligned}$$

L'application directe du théorème 12 donne

$$d(\mathcal{X}_1, \mathcal{X}_1^{(n)}) \leq \frac{\|R\|}{l^2}.$$

Pour clore la démonstration, il reste à majorer  $\|R\|$ . D'après (3.26) et (3.27), on obtient

$$\begin{aligned} \|R\| &= \|Y_1 K_1 \Gamma K_2 \Theta^T + Y_2 (K_2 \Gamma^T K_1 + K_2 K_2^T \Theta^T)\| \\ &\leq \|K_2\| (2 \|K_1\| + \|K_2\|). \end{aligned}$$

Or on a

$$\|K_1\| \leq \kappa_2(X) \frac{\kappa_2(\Phi_1)}{1 - 4\alpha}$$

et

$$\|K_2\| \leq \kappa_2(X) \frac{\omega \gamma^{2^n}}{\alpha - \omega \gamma^{2^n}}.$$

donc  $\forall n \geq n_1$ ,

$$\|R\| \leq \kappa_2^2(X) \left[ 2 \frac{\kappa_2(\Phi_1)}{1 - 4\alpha} + 1 \right] \frac{\omega\gamma^{2^n}}{\alpha - \omega\gamma^{2^n}}.$$

□

**Remarque 7** La présence du carré de  $\kappa_2(X)$ , le conditionnement spectral de  $W$ , provient du fait que cette majoration a été obtenue via la matrice  $X_1^{(n)}(X_1^{(n)})^T$ .

### 3.4.2 Approximation de $\mathbb{P}_1$

Le théorème 11 permet d'estimer la qualité de l'approximation  $\tilde{X}_1^{(n)}$  construite en utilisant la décomposition en valeurs singulières de  $X_1^{(n)}$ . Cette approximation est utilisée pour construire le projecteur  $\tilde{\mathbb{P}}_1^{(n)}$  une approximation de  $\mathbb{P}_1$  intervenant dans le calcul d'approximations de  $\mathcal{X}_0$ ,  $\mathcal{X}_\infty$ ,  $\mathbb{P}_0$  et  $\mathbb{P}_\infty$ .

Le théorème suivant permet de majorer l'erreur obtenue en approchant le projecteur  $\mathbb{P}_1$  par le projecteur  $\tilde{\mathbb{P}}_1^{(n)}$  en fonction de la distance  $d(\mathcal{X}_1, \tilde{\mathcal{X}}_1^{(n)})$ .

**Théorème 13** Soit  $Y_1$  une matrice dont les colonnes forment une base orthonormée de  $\mathcal{X}_1$  et soit  $\eta^{(n)}$  défini par

$$\eta^{(n)} = \left( 3\|J\| \left\| (Y_1^T J Y_1)^{-1} \right\| + 1 \right)^{-1} - d(\mathcal{X}_1, \tilde{\mathcal{X}}_1^{(n)}).$$

Si  $\eta^{(n)} > 0$  alors

$$\left\| \mathbb{P}_1 - \tilde{\mathbb{P}}_1^{(n)} \right\| \leq \frac{K_P}{\eta^{(n)}} d(\mathcal{X}_1, \tilde{\mathcal{X}}_1^{(n)})$$

avec  $K_P$  une constante indépendante de  $n$  donnée par

$$K_P = \left( 2\sqrt{2} + \left( \frac{2\kappa_2(Y_1^T J Y_1)}{3\|J\| \left\| (Y_1^T J Y_1)^{-1} \right\| + 1} + 1 \right) \right) \|J\| \left\| (Y_1^T J Y_1)^{-1} \right\|.$$

#### Démonstration

Soit  $\tilde{Y}_1^{(n)}$  une matrice dont les colonnes forment une base orthonormée de  $\tilde{\mathcal{X}}_1^{(n)}$ . On définit le projecteur  $\tilde{\mathbb{P}}_1^{(n)}$  par

$$\tilde{\mathbb{P}}_1^{(n)} = \tilde{Y}_1^{(n)} \left( (\tilde{Y}_1^{(n)})^T J \tilde{Y}_1^{(n)} \right)^{-1} (\tilde{Y}_1^{(n)})^T J.$$



L'hypothèse sur  $\eta^{(n)} > 0$  implique que  $d(\mathcal{X}_1, \tilde{\mathcal{X}}_1^{(n)}) < 1$  donc d'après la proposition 5, il existe deux matrices  $S$  et  $C$  symétriques, semi-définie positive et définie positive respectivement et une matrice  $Y_1$  dont les colonnes forment une base orthonormée de  $\mathcal{X}_1$  telles que

$$\tilde{Y}_1^{(n)} = Y_1 C + Z S \quad \text{avec} \quad \begin{cases} \|S\| = d(\mathcal{X}_1, \tilde{\mathcal{X}}_1^{(n)}), \\ \lambda_{\min}(C) = \sqrt{1 - d^2(\mathcal{X}_1, \tilde{\mathcal{X}}_1^{(n)})}, \\ Z^T Y_1 = 0, \\ Z^T Z = I \end{cases}$$

On en déduit les résultats suivants :

1. Résultat 1

$$\|Y_1 - \tilde{Y}_1^{(n)}\| = \sqrt{\|I - C\|^2 + \|S\|^2} \leq \sqrt{2} d(\mathcal{X}_1, \tilde{\mathcal{X}}_1^{(n)}),$$

2. Résultat 2

$$\begin{aligned} \left\| (\tilde{Y}_1^{(n)})^T J \tilde{Y}_1^{(n)} - Y_1^T J Y_1 \right\| &\leq (1 + \|C\|) \|Y_1^T J Y_1\| \|I - C\| + 3\|J\| \|S\| \\ &\leq (2 \|Y_1^T J Y_1\| + 3\|J\|) d(\mathcal{X}_1, \tilde{\mathcal{X}}_1^{(n)}), \end{aligned}$$

3. Résultat 3

$$\begin{aligned} \sigma_{\min} \left( (\tilde{Y}_1^{(n)})^T J \tilde{Y}_1^{(n)} \right) &\geq \sigma_{\min}^2(C) \sigma_{\min}(Y_1^T J Y_1) - 3\|J\| \|S\| \\ &\geq \sigma_{\min}(Y_1^T J Y_1) - ((\sigma_{\min}(Y_1^T J Y_1)) + 3\|J\|) d(\mathcal{X}_1, \tilde{\mathcal{X}}_1^{(n)}) \end{aligned}$$

et si  $d(\mathcal{X}_1, \tilde{\mathcal{X}}_1^{(n)}) < \left( 3\|J\| \left\| (Y_1^T J Y_1)^{-1} \right\| + 1 \right)^{-1}$  alors

$$\begin{aligned} \left\| \left( (\tilde{Y}_1^{(n)})^T J \tilde{Y}_1^{(n)} \right)^{-1} \right\| &\leq \frac{\left\| (Y_1^T J Y_1)^{-1} \right\|}{1 - \left( 3\|J\| \left\| (Y_1^T J Y_1)^{-1} \right\| + 1 \right) d(\mathcal{X}_1, \tilde{\mathcal{X}}_1^{(n)})} \\ &\leq \frac{\left\| (Y_1^T J Y_1)^{-1} \right\|}{3\|J\| \left\| (Y_1^T J Y_1)^{-1} \right\| + 1} \frac{1}{\eta^{(n)}}. \end{aligned}$$

En se basant sur ces résultats, on montre que

$$\begin{aligned} \left\| \mathbb{P}_1 - \tilde{\mathbb{P}}_1^{(n)} \right\| &\leq \left\| Y_1 - \tilde{Y}_1^{(n)} \right\| \left\| (Y_1^T J Y_1)^{-1} Y_1^T J \right\| \\ &\quad + \left\| \tilde{Y}_1^{(n)} (Y_1^T J Y_1)^{-1} \right\| \left\| (Y_1 - \tilde{Y}_1^{(n)})^T J \right\| \\ &\quad + \left\| \tilde{Y}_1^{(n)} \left( (Y_1^T J Y_1)^{-1} - (\tilde{Y}_1^{(n)})^T J \tilde{Y}_1^{(n)} \right)^{-1} (\tilde{Y}_1^{(n)})^T J \right\| \end{aligned}$$

Donc on a

$$\begin{aligned} \left\| \mathbb{P}_1 - \tilde{\mathbb{P}}_1^{(n)} \right\| &\leq 2\sqrt{2} \|J\| \left\| (Y_1^T J Y_1)^{-1} \right\| d(\mathcal{X}_1, \tilde{\mathcal{X}}_1^{(n)}) \\ &\quad + \left\| ((\tilde{Y}_1^{(n)})^T J \tilde{Y}_1^{(n)})^{-1} \right\| \left\| (\tilde{Y}_1^{(n)})^T J \tilde{Y}_1^{(n)} - Y_1^T J Y_1 \right\| \left\| (Y_1^T J Y_1)^{-1} \right\|. \end{aligned}$$

soit

$$\begin{aligned} \left\| \mathbb{P}_1 - \tilde{\mathbb{P}}_1^{(n)} \right\| &\leq \left( 2\sqrt{2} + \left\| (Y_1^T J Y_1)^{-1} \right\| \frac{2 \|Y_1^T J Y_1\| + 3 \|J\|}{3 \|J\| \left\| (Y_1^T J Y_1)^{-1} \right\| + 1} \frac{1}{\eta^{(n)}} \right) \\ &\quad \times \|J\| \left\| (Y_1^T J Y_1)^{-1} \right\| d(\mathcal{X}_1, \tilde{\mathcal{X}}_1^{(n)}) \end{aligned}$$

ce qui implique

$$\begin{aligned} &\leq \left( 2\sqrt{2} + \left( \frac{2\kappa_2 (Y_1^T J Y_1) - 1}{3 \|J\| \left\| (Y_1^T J Y_1)^{-1} \right\| + 1} + 1 \right) \frac{1}{\eta^{(n)}} \right) \\ &\quad \times \|J\| \left\| (Y_1^T J Y_1)^{-1} \right\| d(\mathcal{X}_1, \tilde{\mathcal{X}}_1^{(n)}). \end{aligned}$$

La démonstration se termine en remarquant que  $0 < \eta^{(n)} < 1$  car  $\eta^{(n)} > 0$  par hypothèse et

$$3 \|J\| \left\| (Y_1^T J Y_1)^{-1} \right\| + 1 > 1$$

donc

$$\eta^{(n)} = \left( 3 \|J\| \left\| (Y_1^T J Y_1)^{-1} \right\| + 1 \right)^{-1} - d(\mathcal{X}_1, \tilde{\mathcal{X}}_1^{(n)}) < 1$$

car  $d(\mathcal{X}_1, \tilde{\mathcal{X}}_1^{(n)}) \geq 0$ .  $\square$

**Remarque 8** *Les majorations utilisées sont loin d'être optimales cependant elles permettent d'obtenir l'ordre de précision souhaité tout en simplifiant les différentes expressions.*

Les théorèmes 11 et 13 montrent que pour  $n$  suffisamment grand, les vitesses de convergence de  $\tilde{\mathcal{X}}_1^{(n)}$  vers  $\mathcal{X}_1$  et de  $\tilde{\mathbb{P}}_1^{(n)}$  vers  $\mathbb{P}_1$  sont, grossièrement, de la forme  $\gamma^{2^n}$  avec  $0 < \gamma < 1$ . Autrement dit, ces suites convergent de manière supergéométrique (ou hypergéométrique).

### 3.5 Approximations des autres projecteurs

La proposition suivante justifie l'utilisation de  $\tilde{\mathbb{P}}_1^{(n)}$  pour extraire  $\mathbb{P}_0$  et  $\mathbb{P}_\infty$  de  $X_0^{(n)}$  et  $X_\infty^{(n)}$ .

**Proposition 11** *Pour  $n \geq n_1$ , on a*

$$\left\| (I - \mathbb{P}_1) X_0^{(n)} - \frac{1}{\alpha} \mathbb{P}_0 \right\| \leq \kappa_2(X) \left( \frac{\omega}{\alpha - \omega \gamma^{2^n}} \right) \gamma^{2^n},$$

et

$$\left\| (I - \mathbb{P}_1) X_\infty^{(n)} - \frac{1}{\alpha} \mathbb{P}_\infty \right\| \leq \kappa_2(X) \left( \frac{\omega}{\alpha - \omega \gamma^{2^n}} \right) \gamma^{2^n}.$$

**Remarque 9** *On a*

$$(I - \mathbb{P}_1) X_0^{(n)} \xrightarrow{n \rightarrow \infty} \frac{1}{\alpha} \mathbb{P}_0,$$

et

$$(I - \mathbb{P}_1) X_\infty^{(n)} \xrightarrow{n \rightarrow \infty} \frac{1}{\alpha} \mathbb{P}_\infty.$$

#### Démonstration

Pour démontrer la première inégalité, on déduit, des propriétés de  $\mathbb{P}_1$ , que

$$\begin{aligned} (I - \mathbb{P}_1) X_0^{(n)} &= X \begin{pmatrix} W_0^{-2^n} L_0^{(n)} & & \\ & 0 & \\ & & W_\infty^{-2^n} L_\infty^{(n)} \end{pmatrix} X^{-1} \\ &= X_0 \begin{pmatrix} W_0^{-2^n} L_0^{(n)} & 0 & 0 \end{pmatrix} X^{-1} + X_\infty \begin{pmatrix} 0 & 0 & W_\infty^{-2^n} L_\infty^{(n)} \end{pmatrix} X^{-1}. \end{aligned}$$

Le lemme 3 implique que, pour  $n \geq n_1 \geq \max\{n_0, n_\infty\}$ , on a

$$\begin{aligned} \left\| (I - \mathbb{P}_1) X_0^{(n)} - \frac{1}{\alpha} \mathbb{P}_0 \right\| &\leq \kappa_2(X) \max \left\{ \left\| W_0^{-2n} L_0^{(n)} - \frac{1}{\alpha} I \right\|, \left\| W_\infty^{-2n} L_\infty^{(n)} \right\| \right\} \\ &\leq \kappa_2(X) \left( \frac{\omega}{\alpha - \omega \gamma^{2n}} \right) \gamma^{2n} \end{aligned}$$

car  $\mathbb{P}_1 = X_1 \begin{pmatrix} 0 & I & 0 \end{pmatrix} X^{-1}$  et  $\alpha - \alpha \omega \gamma^{2n} \geq \alpha - \omega \gamma^{2n}$ .

La démonstration est identique pour l'autre résultat.  $\square$

**Corollaire 1** *Sous les hypothèses de la proposition 11, on a*

$$\text{eng} \left( (I - \mathbb{P}_1) X_0^{(n)} \right) \xrightarrow[n \rightarrow \infty]{(sse)} \mathcal{X}_0$$

et

$$\text{eng} \left( (I - \mathbb{P}_1) X_\infty^{(n)} \right) \xrightarrow[n \rightarrow \infty]{(sse)} \mathcal{X}_\infty.$$

### Démonstration

On a

$$(I - \mathbb{P}_1) X_0^{(n)} = X_0 W_0^{-2n} L_0^{(n)} Y_0 + X_\infty W_\infty^{-2n} L_\infty^{(n)} Y_\infty.$$

Le premier résultat se montre directement à partir des lemmes 3, 4 et la proposition 9 en posant

$$\tilde{A}_n = X_0 \begin{pmatrix} W_0^{-2n} L_0^{(n)} & 0 & 0 \end{pmatrix} X^{-1} \text{ et } \hat{A}_n = X_\infty \begin{pmatrix} 0 & 0 & W_\infty^{-2n} L_\infty^{(n)} \end{pmatrix} X^{-1}.$$

L'autre démonstration est identique.  $\square$

Ces résultats montrent qu'il est possible de déduire de  $X_0^{(n)}$  et  $X_\infty^{(n)}$ , des approximations de  $\mathbb{P}_0$  et  $\mathbb{P}_\infty$  à condition de disposer du projecteur  $\mathbb{P}_1$ . Or, seule une approximation  $\tilde{\mathbb{P}}_1^{(n)}$  de  $\mathbb{P}_1$  a été calculée. On note  $\tilde{\mathbb{P}}_0^{(n)}$  et  $\tilde{\mathbb{P}}_\infty^{(n)}$  les quantités définies par

$$\tilde{\mathbb{P}}_0^{(n)} = \alpha \left( I - \tilde{\mathbb{P}}_1^{(n)} \right) X_0^{(n)} \quad \text{et} \quad \tilde{\mathbb{P}}_\infty^{(n)} = \alpha \left( I - \tilde{\mathbb{P}}_1^{(n)} \right) X_\infty^{(n)}.$$

Le théorème suivant permet de prendre en compte l'erreur introduite en remplaçant  $\mathbb{P}_1$  par  $\tilde{\mathbb{P}}_1^{(n)}$ .

**Théorème 14** *Sous les hypothèses (3.22) et (3.23), si  $n \geq n_1$  alors*

$$\left\| \tilde{\mathbb{P}}_0^{(n)} - \mathbb{P}_0 \right\| \leq \alpha \kappa_2(X) \left[ \frac{K_p K_X}{\eta^{(n)}} \left( \frac{3\omega}{\alpha - \omega \gamma^{2n}} + \frac{\kappa_2(\phi_1)}{1 - 4\alpha} \right) + 1 \right] \frac{\omega \gamma^{2n}}{\alpha - \omega \gamma^{2n}},$$

et

$$\left\| \tilde{\mathbb{P}}_{\infty}^{(n)} - \mathbb{P}_{\infty} \right\| \leq \alpha \kappa_2(X) \left[ \frac{K_p K_X}{\eta^{(n)}} \left( \frac{3\omega}{\alpha - \omega\gamma^{2^n}} + \frac{\kappa_2(\phi_1)}{1 - 4\alpha} \right) + 1 \right] \frac{\omega\gamma^{2^n}}{\alpha - \omega\gamma^{2^n}}.$$

### Démonstration

On a les majorations suivantes

$$\begin{aligned} \left\| \tilde{\mathbb{P}}_0^{(n)} - \mathbb{P}_0 \right\| &\leq \left\| \alpha(I - \tilde{\mathbb{P}}_1^{(n)})X_0^{(n)} - \alpha(I - \mathbb{P}_1)X_0^{(n)} + \alpha(I - \mathbb{P}_1)X_0^{(n)} - \mathbb{P}_0 \right\| \\ &\leq \alpha \left\| \tilde{\mathbb{P}}_1^{(n)} - \mathbb{P}_1 \right\| \left\| X_0^{(n)} \right\| + \alpha \left\| (I - \mathbb{P}_1)X_0^{(n)} - \frac{1}{\alpha}\mathbb{P}_0 \right\|. \end{aligned}$$

En utilisant les résultats des lemmes 3 et 4, on montre que, pour  $n \geq n_1$ ,

$$\begin{aligned} \left\| X_0^{(n)} \right\| &\leq \kappa_2(X) \left( \frac{1}{\alpha} + \frac{1}{\alpha} \frac{\omega_0\gamma_0^{2^n}}{1 - \omega_0\gamma_0^{2^n}} + \frac{\kappa_2(\phi_1)}{1 - 4\alpha} + \frac{\omega_{\infty}\gamma_{\infty}^{2^{n+1}}}{\alpha - \omega_{\infty}\gamma_{\infty}^{2^n}} \right) \\ &\leq \kappa_2(X) \left( 3 \frac{\omega}{\alpha - \omega\gamma^{2^n}} + \frac{\kappa_2(\phi_1)}{1 - 4\alpha} \right). \end{aligned}$$

La démonstration se termine en utilisant la proposition 11 et les théorèmes 11 et 13.  $\square$

**Corollaire 2** *Sous les hypothèses du théorème 14, on a*

$$\text{eng} \left( \tilde{\mathbb{P}}_0^{(n)} \right) \xrightarrow[n \rightarrow \infty]{(sse)} \mathcal{X}_0$$

et

$$\text{eng} \left( \tilde{\mathbb{P}}_{\infty}^{(n)} \right) \xrightarrow[n \rightarrow \infty]{(sse)} \mathcal{X}_{\infty}.$$

### Démonstration

Le théorème 14 nous donne la convergence de  $\tilde{\mathbb{P}}_0^{(n)}$  et  $\tilde{\mathbb{P}}_{\infty}^{(n)}$  vers les matrices  $\mathbb{P}_0$  et  $\mathbb{P}_{\infty}$  respectivement lorsque  $n$  tend vers l'infini. Or,  $\text{eng}(\mathbb{P}_0) = \mathcal{X}_0$  et  $\text{eng}(\mathbb{P}_{\infty}) = \mathcal{X}_{\infty}$ . Alors, d'après la proposition 6,

$$\text{eng} \left( \tilde{\mathbb{P}}_0^{(n)} \right) \xrightarrow[n \rightarrow \infty]{(sse)} \mathcal{X}_0$$

et

$$\text{eng} \left( \tilde{\mathbb{P}}_{\infty}^{(n)} \right) \xrightarrow[n \rightarrow \infty]{(sse)} \mathcal{X}_{\infty}.$$

$\square$

**Remarque 10** *Il est important de noter que si  $\tilde{\mathbb{P}}_0^{(n)}$  et  $\tilde{\mathbb{P}}_\infty^{(n)}$  convergent vers des projecteurs, ce ne sont a priori pas des projecteurs. Il est parfois préférable, suivant l'utilisation faite de  $\tilde{\mathbb{P}}_0^{(n)}$  et  $\tilde{\mathbb{P}}_\infty^{(n)}$  par la suite, de disposer d'approximations de projecteurs qui soient elles-même des projecteurs. La construction de projecteurs à partir de  $\tilde{\mathbb{P}}_0^{(n)}$  et  $\tilde{\mathbb{P}}_\infty^{(n)}$  ainsi que leur qualité sont discutées dans la section suivante car ce point ne concerne pas directement l'algorithme de trichotomie.*

## 3.6 Mise en œuvre informatique

L'implantation de l'algorithme de trichotomie est délicate. Les principaux problèmes sont le calcul effectif des suites  $(X_0^{(n)})_{n \geq 0}$ ,  $(X_1^{(n)})_{n \geq 0}$  et  $(X_\infty^{(n)})_{n \geq 0}$  et ainsi que celui des approximations des projecteurs  $\mathbb{P}_0$ ,  $\mathbb{P}_1$  et  $\mathbb{P}_\infty$ . Ces points font l'objet d'une étude particulière. L'algorithme de la méthode de trichotomie conclut cette section.

### 3.6.1 Calcul de $X_0^{(n)}$ , $X_1^{(n)}$ et $X_\infty^{(n)}$

Il est possible de déduire des expressions (3.2), (3.3) et (3.4), une méthode simple permettant de calculer  $X_0^{(n)}$ ,  $X_1^{(n)}$  et  $X_\infty^{(n)}$ . Cette méthode est la suivante

#### Méthode 1

1. Poser  $V_0 = W$ .
2. Pour  $j = 1$  à  $n$ , calculer
  - $V_j = V_{j-1}^2$ .
3. Calculer  $T_n = V_n^{-1}$ .
4. Calculer
  - $G_n = (\beta I + \alpha (V_n + T_n))^{-1}$
  - $X_0^{(n)} = T_n G_n$ ,
  - $X_1^{(n)} = G_n$ ,
  - $X_\infty^{(n)} = V_n G_n$ .

Cette méthode est directement basée sur le calcul de  $W^{2^n}$ . Or il est connu que le calcul des puissances d'une matrice est une opération numériquement instable. En effet, le conditionnement de la matrice  $V_n$  risque d'augmenter très rapidement avec  $n$ , rendant difficile un calcul précis de  $T_n$ . Pour éviter de devoir inverser  $V_n$ , une solution consiste à calculer  $T_n$  indépendamment de  $V_n$ . Cette option est utilisée dans la méthode suivante.

### Méthode 2

1. Poser  $V_0 = W$  et  $T_0 = W^{-1}$ .
2. Pour  $j = 1$  à  $n$ , calculer
  - $V_j = V_{j-1}^2$ .
  - $T_j = T_{j-1}^2$ .
3. Calculer
  - $G_n = (\beta I + \alpha (V_n + T_n))^{-1}$
  - $X_0^{(n)} = T_n G_n$ ,
  - $X_1^{(n)} = G_n$ ,
  - $X_\infty^{(n)} = V_n G_n$ .

Cependant certains problèmes numériques persistent notamment lors du calcul de  $G_n$ . En effet, les hypothèses impliquant que les normes des matrices  $X_0^{(n)}$ ,  $X_1^{(n)}$  et  $X_\infty^{(n)}$  sont bornées, ne sont pas suffisantes pour montrer que la norme de  $W^{2^n}$  est bornée ou autrement dit que  $W$  est stable. Il est donc possible, voir probable que les méthodes 1 et 2 rencontrent de sérieux problèmes numériques. De nombreuses expériences numériques ont montré que ces craintes sont justifiées.

Afin d'effectuer le calcul de  $X_0^{(n)}$ ,  $X_1^{(n)}$  et  $X_\infty^{(n)}$  de manière plus stable, nous utilisons une nouvelle méthode basée sur la résolution d'un système linéaire. Si le coût de calcul de cette méthode est plus important, elle permet de fiabiliser fortement les résultats obtenus.

Cette méthode utilise les relations entre les matrices  $X_0^{(n-1)}$ ,  $X_1^{(n-1)}$  et  $X_\infty^{(n-1)}$  et  $X_0^{(n)}$ ,  $X_1^{(n)}$  et  $X_\infty^{(n)}$  de la proposition suivante.

**Proposition 12** *Si  $\alpha$  et  $\beta$  sont choisis tels que, pour tout  $n \geq 0$ , la matrice  $\beta I + \alpha (W^{2^n} + W^{-2^n})$  est non singulière alors  $X_0^{(1)}$ ,  $X_1^{(1)}$  et  $X_\infty^{(1)}$  forment l'unique solution du système*

$$\begin{pmatrix} JW & -W^T J & 0 \\ 0 & JW & -W^T J \\ \alpha I & \beta I & \alpha I \end{pmatrix} \begin{pmatrix} X_0^{(1)} \\ X_1^{(1)} \\ X_\infty^{(1)} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ I \end{pmatrix} \quad (3.28)$$

et pour  $n \geq 2$ ,

$$\begin{aligned} X_0^{(n)} &= X_0^{(n-1)} K_0^{(n)}, \\ X_1^{(n)} &= X_1^{(n-1)} K_1^{(n)}, \\ X_\infty^{(n)} &= X_\infty^{(n-1)} K_\infty^{(n)} \end{aligned}$$

où  $K_0^{(n)}$ ,  $K_1^{(n)}$  et  $K_\infty^{(n)}$  forment la solution du système linéaire

$$\begin{pmatrix} \left(X_1^{(n-1)}\right)^T J & -\left(X_\infty^{(n-1)}\right)^T J & 0 \\ 0 & \left(X_0^{(n-1)}\right)^T J & -\left(X_1^{(n-1)}\right)^T J \\ \alpha X_0^{(n-1)} & \beta X_1^{(n-1)} & \alpha X_\infty^{(n-1)} \end{pmatrix} \begin{pmatrix} K_0^{(n)} \\ K_1^{(n)} \\ K_\infty^{(n)} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ I \end{pmatrix}. \quad (3.29)$$

### Démonstration

Sachant que

$$\begin{aligned} X_0^{(1)} &= W^{-2} (\beta I + \alpha (W^2 + W^{-2}))^{-1}, \\ X_1^{(1)} &= (\beta I + \alpha (W^2 + W^{-2}))^{-1}, \\ X_\infty^{(1)} &= W^2 (\beta I + \alpha (W^2 + W^{-2}))^{-1}, \end{aligned}$$

on a les relations

$$\begin{aligned} X_0^{(1)} &= W^{-2} X_1^{(1)} \\ X_1^{(1)} &= W^{-2} X_\infty^{(1)} \\ \alpha X_0^{(1)} + \beta X_1^{(1)} + \alpha X_\infty^{(1)} &= I. \end{aligned}$$

On déduit de la  $J$ -symplecticité de  $W$  que  $W^T J = J W^{-1}$ . On a donc les relations

$$\begin{aligned} J W X_0^{(1)} &= J W^{-1} X_1^{(1)} = W^T J X_1^{(1)} \\ J W X_1^{(1)} &= J W^{-1} X_\infty^{(1)} = W^T J X_\infty^{(1)} \\ \alpha X_0^{(1)} + \beta X_1^{(1)} + \alpha X_\infty^{(1)} &= I. \end{aligned}$$

Autrement dit  $X_0^{(1)}$ ,  $X_1^{(1)}$  et  $X_\infty^{(1)}$  forment une solution du système (3.28).

D'autre part, si  $A$ ,  $B$  et  $C$  sont tels que

$$\begin{pmatrix} J W & -W^T J & 0 \\ 0 & J W & -W^T J \\ \alpha I & \beta I & \alpha I \end{pmatrix} \begin{pmatrix} A \\ B \\ C \end{pmatrix} = 0$$

alors  $J$  étant non singulière, on a les relations

$$\begin{aligned} A &= W^{-2} B \\ B &= W^{-2} C \\ \alpha A + \beta B + \alpha C &= 0. \end{aligned}$$



On en déduit que  $B$  vérifie

$$(\beta I + \alpha (W^{2^n} + W^{-2^n})) B = 0.$$

La matrice  $\beta I + \alpha (W^{2^n} + W^{-2^n})$  étant supposée non singulière, la matrice  $B$  est nulle et par conséquent les matrices  $A$  et  $C$  le sont également. Autrement dit,  $X_0^{(1)}$ ,  $X_1^{(1)}$  et  $X_\infty^{(1)}$  forment l'unique solution du système (3.28).

Pour  $n \geq 2$ , on peut écrire avec  $p = n$  et  $p = n - 1$ ,

$$\begin{aligned} X_0^{(p)} &= W^{-2^p} L^{(p)}, \\ X_1^{(p)} &= L^{(p)}, \\ X_\infty^{(p)} &= W^{2^p} L^{(p)}, \end{aligned}$$

avec  $L^{(p)} = (\beta I + \alpha (W^{2^p} + W^{-2^p}))^{-1}$ . De plus, il est clair que  $W$  et  $L^{(p)}$  commutent.

Nous allons montrer que les matrices  $K_0^{(n)}$ ,  $K_1^{(n)}$  et  $K_\infty^{(n)}$  définies par

$$\begin{aligned} K_0^{(n)} &\equiv X_0^{(n)} \left( X_0^{(n-1)} \right)^{-1} = W^{-2^{n-1}} L^{(n)} \left( L^{(n-1)} \right)^{-1}, \\ K_1^{(n)} &\equiv X_1^{(n)} \left( X_1^{(n-1)} \right)^{-1} = L^{(n)} \left( L^{(n-1)} \right)^{-1}, \\ K_\infty^{(n)} &\equiv X_\infty^{(n)} \left( X_\infty^{(n-1)} \right)^{-1} = W^{2^{n-1}} L^{(n)} \left( L^{(n-1)} \right)^{-1} \end{aligned}$$

forment une solution du système (3.29).

En effet, on montre que

$$\begin{aligned} &\left( X_1^{(n-1)} \right)^T J K_0^{(n)} - \left( X_\infty^{(n-1)} \right)^T J K_1^{(n)} \\ &= \left( L^{(n-1)} \right)^T \left( J W^{-2^{n-1}} - \left( W^{2^{n-1}} \right)^T J \right) L^{(n)} \left( L^{(n-1)} \right)^{-1} \\ &= 0 \end{aligned}$$

car on déduit de la  $J$ -symplecticité de  $W$  que

$$J W^{-2^{n-1}} - \left( W^{2^{n-1}} \right)^T J = \left( J - \left( W^{2^{n-1}} \right)^T J W^{2^{n-1}} \right) W^{-2^{n-1}} = 0.$$

On montre de même que

$$\begin{aligned} &\left( X_0^{(n-1)} \right)^T J K_1^{(n)} - \left( X_1^{(n-1)} \right)^T J K_\infty^{(n)} \\ &= \left( L^{(n-1)} \right)^T \left( \left( W^{-2^{n-1}} \right)^T J - J W^{2^{n-1}} \right) L^{(n)} \left( L^{(n-1)} \right)^{-1} \\ &= 0 \end{aligned}$$

La dernière relation s'écrit

$$\begin{aligned} \alpha X_0^{(n-1)} K_0^{(n)} + \beta X_1^{(n-1)} K_1^{(n)} + \alpha X_\infty^{(n-1)} K_\infty^{(n)} &= \\ &= L^{(n-1)} (\alpha W^{-2^n} + \beta I + \alpha W^{2^n}) L^{(n)} (L^{(n-1)})^{-1} \\ &= L^{(n-1)} (L^{(n)})^{-1} L^{(n)} (L^{(n-1)})^{-1} = I. \end{aligned}$$

Pour clore la démonstration, il reste à montrer que  $K_0^{(n)}$ ,  $K_1^{(n)}$  et  $K_\infty^{(n)}$  forment l'unique solution du système (3.29).

On suppose que les vecteurs  $v_0$ ,  $v_1$  et  $v_\infty$  vérifient

$$\begin{pmatrix} (X_1^{(n-1)})^T J & - (X_\infty^{(n-1)})^T J & 0 \\ 0 & (X_0^{(n-1)})^T J & - (X_1^{(n-1)})^T J \\ \alpha X_0^{(n-1)} & \beta X_1^{(n-1)} & \alpha X_\infty^{(n-1)} \end{pmatrix} \begin{pmatrix} v_0 \\ v_1 \\ v_\infty \end{pmatrix} = 0.$$

On déduit, des deux premières lignes, les relations

$$Jv_0 = (W^{2^{n-1}})^T Jv_1 = JW^{-2^{n-1}} v_1$$

et

$$Jv_\infty = (W^{-2^{n-1}})^T Jv_1 = JW^{2^{n-1}} v_1$$

en utilisant la  $J$ -symplecticité de  $W$ . La matrice  $J$  étant non singulière, ces relations peuvent donc s'écrire de manière équivalente

$$v_0 = W^{-2^{n-1}} v_1 \text{ et } v_\infty = W^{2^{n-1}} v_1.$$

Elles permettent en association avec la troisième ligne du système linéaire de déduire que le vecteur  $v_1$  est nul. en effet, on a

$$\begin{aligned} \alpha X_0^{(n-1)} v_0 + \beta X_1^{(n-1)} v_1 + \alpha X_\infty^{(n-1)} v_\infty &= \\ &= L^{(n-1)} (\alpha W^{-2^{n-1}} v_0 + \beta v_1 + \alpha W^{2^{n-1}} v_\infty) \\ &= L^{(n-1)} (\alpha W^{-2^n} + \beta I + \alpha W^{2^n}) v_1 \\ &= L^{(n-1)} (L^{(n)})^{-1} v_1 \end{aligned}$$

et donc la condition

$$\alpha X_0^{(n-1)} v_0 + \beta X_1^{(n-1)} v_1 + \alpha X_\infty^{(n-1)} v_\infty = 0$$

implique que  $v_1 = 0$  car  $L^{(n-1)}$  est non singulière. Par conséquent, il est clair que  $v_0$  et  $v_\infty$  sont également nuls. Autrement dit, le système (3.29) est non singulier.  $\square$

**Remarque 11** *Il est possible de remplacer le système (3.29) par d'autres systèmes qui lui sont similaires.*

En se basant sur la proposition 12, on obtient une nouvelle méthode.

### Méthode 3

#### 1. Résoudre

$$\begin{pmatrix} JW & -W^T J & 0 \\ 0 & JW & -W^T J \\ \alpha I & \beta I & \alpha I \end{pmatrix} \begin{pmatrix} K_0^{(1)} \\ K_1^{(1)} \\ K_\infty^{(1)} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ I \end{pmatrix}.$$

#### 2. Poser

- $X_0^{(1)} = K_0^{(1)},$
- $X_1^{(1)} = K_1^{(1)},$
- $X_\infty^{(1)} = K_\infty^{(1)}.$

#### 3. Pour $j = 2$ à $n$ ,

##### - Résoudre

$$\begin{pmatrix} \left(X_1^{(j-1)}\right)^T J & -\left(X_\infty^{(j-1)}\right)^T J & 0 \\ 0 & \left(X_0^{(j-1)}\right)^T J & -\left(X_1^{(j-1)}\right)^T J \\ \alpha X_0^{(j-1)} & \beta X_1^{(j-1)} & \alpha X_\infty^{(j-1)} \end{pmatrix} \begin{pmatrix} K_0^{(j)} \\ K_1^{(j)} \\ K_\infty^{(j)} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ I \end{pmatrix}.$$

##### - Poser

- .  $X_0^{(j)} = X_0^{(j-1)} K_0^{(j)},$
- .  $X_1^{(j)} = X_1^{(j-1)} K_1^{(j)},$
- .  $X_\infty^{(j)} = X_\infty^{(j-1)} K_\infty^{(j)}.$

Le coût de calcul de la méthode 3 est évidemment plus élevé que ceux de la méthode 1 et de la méthode 2. Les tableaux 3.1, 3.2 et 3.3 donnent les estimations du coût de calcul de chaque méthode, réalisées de la même manière que celles données dans le Chapitre 2 (i.e. en négligeant les termes inférieurs à  $N^3$ ).

Si la méthode 3 s'est révélée plus stable numériquement que les méthodes 1 et 2, elle reste sujet à des complications numériques lors de la résolution des systèmes linéaires. Cependant ces problèmes surviennent lorsque la méthode de trichotomie a convergé. Ils peuvent donc être évités en utilisant un critère d'arrêt adéquat.

<hr/>		
<u>Itérations</u>	( $n$ fois)	
1 produit matrice matrice		$2(2N)^3$
<u>Post-traitement</u>		
1 inversion de matrice ( $T_n$ )		
– Une factorisation LU		$2/3(2N)^3$
– $2 \times (2N)$ systèmes linéaires triangulaires		$2(2N)^3$
1 inversion de matrice ( $G_n$ )		
– Une factorisation LU		$2/3(2N)^3$
– $2 \times (2N)$ systèmes linéaires triangulaires		$2(2N)^3$
2 produits matrice matrice		$4(2N)^3$
<hr/>		
Total		$(2n + 28/3)(2N)^3$

TAB. 3.1 – Coût des opérations de la méthode 1

<hr/>		
<u>Initialisation</u>		
1 inversion de matrice ( $T_0$ )		
– Une factorisation LU		$2/3(2N)^3$
– $2 \times (2N)$ systèmes linéaires triangulaires		$2(2N)^3$
<u>Itérations</u>	( $n$ fois)	
1 produit matrice matrice		$2(2N)^3$
1 produit matrice matrice		$2(2N)^3$
<u>Post-traitement</u>		
1 inversion de matrice ( $G_n$ )		
– Une factorisation LU		$2/3(2N)^3$
– $2 \times (2N)$ systèmes linéaires triangulaires		$2(2N)^3$
2 produits matrice matrice		$4(2N)^3$
<hr/>		
Total		$(4n + 28/3)(2N)^3$

TAB. 3.2 – Coût des opérations de la méthode 2

---



---

<u>Initialisation</u>		
2 produits matrice matrice		$4(2N)^3$
$2N$ résolutions de systèmes linéaires		
– Une factorisation LU		$18(\times 2N)^3$
– $\times(2N)$ systèmes linéaires triangulaires		$10(2N)^3$
<u>Itérations</u> <span style="float: right;">(<math>n - 1</math> fois)</span>		
2 produits matrice matrice		$4(2N)^3$
$2N$ résolutions de systèmes linéaires		
– Une factorisation LU		$18(2N)^3$
– $2 \times (2N)$ systèmes linéaires triangulaires		$10(2N)^3$
3 produits matrice matrice		$6(2N)^3$
Total $(38n - 6)(2N)^3$		

---



---

TAB. 3.3 – Coût des opérations de la méthode 3

### 3.6.2 Approximations des projecteurs

Dans cette partie, on suppose que l'algorithme de trichotomie a calculé des matrices  $\tilde{X}_0^{(n)}$ ,  $\tilde{\mathbb{P}}_1^{(n)}$  et  $\tilde{X}_\infty^{(n)}$  vérifiant

$$\left\| \alpha \tilde{X}_0^{(n)} - \mathbb{P}_0 \right\| \leq \epsilon, \quad \left\| \tilde{\mathbb{P}}_1^{(n)} - \mathbb{P}_1 \right\| \leq \epsilon \quad \text{et} \quad \left\| \alpha \tilde{X}_\infty^{(n)} - \mathbb{P}_\infty \right\| \leq \epsilon$$

où une majoration de  $\epsilon$  peut être obtenue via les résultats des sections précédentes.

Il est donc possible de stopper l'algorithme de trichotomie à cet instant en prenant  $\tilde{\mathbb{P}}_0^{(n)} \equiv \alpha \tilde{X}_0^{(n)}$ ,  $\tilde{\mathbb{P}}_1^{(n)}$  et  $\tilde{\mathbb{P}}_\infty^{(n)} \equiv \alpha \tilde{X}_\infty^{(n)}$  comme approximations de  $\mathbb{P}_0$ ,  $\mathbb{P}_1$  et  $\mathbb{P}_\infty$  respectivement. Cependant si  $\tilde{\mathbb{P}}_1^{(n)}$  est un projecteur (par construction), cela n'est pas le cas de  $\tilde{\mathbb{P}}_0^{(n)}$  et  $\tilde{\mathbb{P}}_\infty^{(n)}$ . Or il est souvent préférable d'approcher un projecteur par un autre projecteur.

La construction, à partir de  $\tilde{X}_0^{(n)}$ , d'un projecteur  $\hat{\mathbb{P}}_0^{(n)}$  qui soit une approximation de  $\mathbb{P}_0$  peut être réalisée de la manière suivante.

1. On construit une matrice  $\hat{X}_0^{(n)}$  dont les colonnes forment une base orthonormée d'une approximation de  $\mathcal{X}_0$ .
2. On construit une matrice  $\hat{Y}_0^{(n)}$  dont les colonnes forment une base orthonormée d'une approximation de  $\mathcal{Y}_0$  le sous-espace invariant à gauche de  $W$  associé à  $W_0$ .

3. On définit  $\widehat{\mathbb{P}}_0^{(n)}$  par

$$\widehat{\mathbb{P}}_0^{(n)} = \widehat{X}_0^{(n)} \left( \left( \widehat{Y}_0^{(n)} \right)^T \widehat{X}_0^{(n)} \right)^{-1} \left( \widehat{Y}_0^{(n)} \right)^T.$$

Il est clair que la matrice  $\widehat{\mathbb{P}}_0^{(n)}$  est un projecteur. En se basant sur le fait que

$$\mathbb{P}_0 = X_0 (Y_0^T X_0)^{-1} Y_0^T$$

où  $Y_0$  est une base orthonormée de  $\mathcal{Y}_0$ , on remarque que l'analyse d'erreur faite en approchant  $\mathbb{P}_0$  par  $\widehat{\mathbb{P}}_0^{(n)}$  se décompose en trois parties

- l'estimation de la qualité de  $\widehat{X}_0^{(n)}$ ,
- l'estimation de la qualité de  $\widehat{Y}_0^{(n)}$ ,
- la majoration de  $\left\| \mathbb{P}_0 - \widehat{\mathbb{P}}_0^{(n)} \right\|$ .

Le lemme suivant permet de répondre à la première partie.

**Lemme 5** *Si*

$$\epsilon < \frac{1}{\|Y_0^T X_0\|}$$

*alors en utilisant une décomposition en valeurs singulières de  $\widetilde{X}_0^{(n)}$ , il est possible de construire  $\widehat{X}_0^{(n)}$  vérifiant*

$$d\left(\mathcal{X}_0, \widehat{\mathcal{X}}_0^{(n)}\right) \leq \|Y_0^T X_0\| \epsilon < 1$$

où  $\widehat{\mathcal{X}}_0^{(n)}$  est l'image de  $\widehat{X}_0^{(n)}$ .

### Démonstration

Soit la décomposition en valeurs singulières de  $\widetilde{X}_0^{(n)}$  suivante :

$$\alpha \widetilde{X}_0^{(n)} = U \Sigma V^T \quad \text{avec} \quad \begin{cases} U \text{ et } V \text{ orthogonales,} \\ \Sigma = \text{diag}(\sigma_1, \dots, \sigma_{2N}) \\ \text{avec } \sigma_1 \geq \dots \geq \sigma_{2N} \geq 0. \end{cases}$$

On décompose les matrices  $U$ ,  $V$  et  $\Sigma$  sous la forme

$$U = (U_1 \quad U_2), \quad V = (V_1 \quad V_2) \quad \text{et} \quad \Sigma = \begin{pmatrix} \Sigma_1 & \\ & \Sigma_2 \end{pmatrix}$$

avec  $U_1, V_1 \in \mathbb{R}^{2N \times N_0}$  et  $\Sigma_1 \in \mathbb{R}^{N_0 \times N_0}$ .

En répétant la démonstration du théorème 10, le théorème de Schmidt-Mirsky permet de montrer que

$$\|\Sigma_2\| \leq \epsilon$$

car  $\mathbb{P}_0$  est de rang  $N_0$ . On définit  $\hat{X}_0^{(n)}$  de la manière suivante :

$$\hat{X}_0^{(n)} \equiv U_1.$$

On déduit des propriétés de la distance  $d$  que

$$\begin{aligned} d\left(\mathcal{X}_0, \hat{\mathcal{X}}_0^{(n)}\right) &\leq \max_{v_1 \in \mathcal{X}_0 - \{0\}} \min_{v_2 \in \hat{\mathcal{X}}_0^{(n)}} \frac{\|v_1 - v_2\|}{\|v_1\|} \\ &= \max_{v_1 \neq 0} \min_{v_2} \frac{\left\|X_0 (Y_0^T X_0)^{-1} v_1 - \hat{X}_0^{(n)} (Y_0^T X_0)^{-1} v_2\right\|}{\left\|X_0 (Y_0^T X_0)^{-1} v_1\right\|} \\ &= \max_{v_1 \in \mathcal{Y}_0 - \{0\}} \min_{v_2} \frac{\left\|X_0 (Y_0^T X_0)^{-1} Y_0^T v_1 - \hat{X}_0^{(n)} (Y_0^T X_0)^{-1} v_2\right\|}{\left\|X_0 (Y_0^T X_0)^{-1} Y_0^T v_1\right\|} \\ &= \max_{v_1 \in \mathcal{Y}_0 - \{0\}} \min_{v_2} \frac{\left\|\mathbb{P}_0 v_1 - \hat{X}_0^{(n)} (Y_0^T X_0)^{-1} v_2\right\|}{\left\|(Y_0^T X_0)^{-1} v_1\right\|}. \end{aligned}$$

En prenant  $v_2 = (Y_0^T X_0) \Sigma_1 V_1^T v_1$ , on obtient

$$\begin{aligned} d\left(\mathcal{X}_0, \hat{\mathcal{X}}_0^{(n)}\right) &\leq \max_{v_1 \in \mathcal{Y}_0 - \{0\}} \frac{\left\|\mathbb{P}_0 v_1 - \hat{X}_0^{(n)} \Sigma_1 V_1^T v_1\right\|}{\left\|(Y_0^T X_0)^{-1} v_1\right\|} \\ &\leq \left( \max_{v_1 \in \mathcal{Y}_0 - \{0\}} \frac{\|v_1\|}{\left\|(Y_0^T X_0)^{-1} v_1\right\|} \right) \epsilon \\ &\leq \|Y_0^T X_0\| \epsilon. \end{aligned}$$

□

**Remarque 12** Si la condition  $\epsilon < \frac{1}{\|Y_0^T X_0\|}$  n'est pas vérifiée alors le résultat du lemme 5 est sans intérêt car la distance  $d$  est toujours inférieure ou égale à 1.

En procédant de la même manière avec les matrices transposées, on obtient le lemme suivant.

**Lemme 6** *Si*

$$\epsilon < \frac{1}{\|Y_0^T X_0\|}$$

*alors en utilisant une décomposition en valeurs singulières de  $\tilde{X}_0^{(n)}$ , il est possible de construire  $\hat{Y}_0^{(n)}$  vérifiant*

$$d(\mathcal{Y}_0, \hat{\mathcal{Y}}_0^{(n)}) \leq \|Y_0^T X_0\| \epsilon < 1$$

*où  $\hat{\mathcal{Y}}_0^{(n)}$  est l'image de  $\hat{Y}_0^{(n)}$ .*

**Remarque 13** *En se reportant à la démonstration du lemme 5, on constate que  $\hat{Y}_0^{(n)} \equiv V_1$  est un choix vérifiant le lemme 6.*

Le théorème suivant donne la majoration de la troisième et dernière étape.

**Théorème 15** *On suppose que*

$$\epsilon < \frac{1}{5 \kappa_2(Y_0^T X_0)}$$

*et on se donne  $\hat{X}_0^{(n)}$  et  $\hat{Y}_0^{(n)}$  vérifiant respectivement les lemmes 5 et 6. La matrice définie par*

$$\hat{\mathbb{P}}_0^{(n)} = \hat{X}_0^{(n)} \left( (\hat{Y}_0^{(n)})^T \hat{X}_0^{(n)} \right)^{-1} (\hat{Y}_0^{(n)})^T$$

*vérifie*

$$\|\mathbb{P}_0 - \hat{\mathbb{P}}_0^{(n)}\| \leq \kappa_2(Y_0^T X_0) \left( 4 + 5 \frac{\kappa_2(Y_0^T X_0)}{1 - 5 \kappa_2(Y_0^T X_0) \epsilon} \right) \epsilon.$$

**Démonstration**

Il est clair que la condition sur  $\epsilon$  implique que  $\epsilon < \frac{1}{\|Y_0^T X_0\|}$ . Il est donc possible d'appliquer les lemmes 5 et 6. On en déduit que  $d(\mathcal{X}_0, \hat{\mathcal{X}}_0^{(n)}) < 1$  et  $d(\mathcal{Y}_0, \hat{\mathcal{Y}}_0^{(n)}) < 1$  et donc on peut utiliser la proposition 5 pour obtenir les décompositions suivantes :

$$\hat{X}_0^{(n)} = X_0 C_x + V_x S_x$$



et

$$\widehat{Y}_0^{(n)} = Y_0 C_y + V_y S_y$$

où

- $S_x$  et  $S_y$  sont symétriques et semi-définies positives,
- $C_x$  et  $C_y$  sont symétriques et définies positives,
- les colonnes de  $V_x$  et  $V_y$  sont orthonormées

et vérifient

$$\begin{aligned} \|S_x\| &= d\left(\mathcal{X}_0, \widehat{\mathcal{X}}_0^{(n)}\right) \leq \|Y_0^T X_0\| \epsilon, \\ \|S_y\| &= d\left(\mathcal{Y}_0, \widehat{\mathcal{Y}}_0^{(n)}\right) \leq \|Y_0^T X_0\| \epsilon, \\ \lambda_{\min}(C_x) &= \sqrt{1 - d^2\left(\mathcal{X}_0, \widehat{\mathcal{X}}_0^{(n)}\right)} \geq \sqrt{1 - \|Y_0^T X_0\|^2 \epsilon^2}, \\ \lambda_{\min}(C_y) &= \sqrt{1 - d^2\left(\mathcal{Y}_0, \widehat{\mathcal{Y}}_0^{(n)}\right)} \geq \sqrt{1 - \|Y_0^T X_0\|^2 \epsilon^2}, \\ X_0^T V_x &= Y_0^T V_y = 0. \end{aligned}$$

On en déduit les inégalités suivantes.

$$\left\|X_0 - \widehat{X}_0^{(n)}\right\| \leq \|I - C_x\| + \|S_x\| \leq 2 \|Y_0^T X_0\| \epsilon$$

et

$$\left\|Y_0 - \widehat{Y}_0^{(n)}\right\| \leq \|I - C_y\| + \|S_y\| \leq 2 \|Y_0^T X_0\| \epsilon.$$

De même, on a

$$\begin{aligned} \left\|Y_0^T X_0 - \left(\widehat{Y}_0^{(n)}\right)^T \widehat{X}_0^{(n)}\right\| &\leq \|Y_0^T X_0 - C_y Y_0^T X_0 C_x\| \\ &\quad + (\|C_y\| \|S_x\| + \|C_x\| \|S_y\| + \|S_x\| \|S_y\|) \|Y_0^T X_0\| \\ &\leq (\|I - C_x\| + \|C_x\| \|I - C_y\|) \|Y_0^T X_0\| + 3 \|Y_0^T X_0\|^2 \epsilon \\ &\leq 5 \|Y_0^T X_0\|^2 \epsilon. \end{aligned}$$

Soit  $A$  et  $B$  deux matrices inversibles telles que

$$\|A^{-1}(B - A)\| = \rho < 1$$

alors

$$\|B^{-1}\| \leq \|A^{-1}\| \|(I + A^{-1}(B - A))^{-1}\| \leq \|A^{-1}\| \frac{1}{1 - \rho}.$$

Or si

$$\epsilon < \frac{1}{5 \kappa_2 (Y_0^T X_0)}$$

alors

$$\begin{aligned} \|(Y_0^T X_0)^{-1} (Y_0^T X_0 - (\hat{Y}_0^{(n)})^T \hat{X}_0^{(n)})\| &\leq \|(Y_0^T X_0)^{-1}\| \|Y_0^T X_0 - (\hat{Y}_0^{(n)})^T \hat{X}_0^{(n)}\| \\ &\leq 5 \|(Y_0^T X_0)^{-1}\| \|Y_0^T X_0\|^2 \epsilon \\ &\leq 5 \kappa_2 (Y_0^T X_0) \|Y_0^T X_0\| \epsilon < 1. \end{aligned}$$

On en déduit que

$$\left\| \left( (\hat{Y}_0^{(n)})^T \hat{X}_0^{(n)} \right)^{-1} \right\| \leq \frac{\|(Y_0^T X_0)^{-1}\|}{1 - 5 \kappa_2 (Y_0^T X_0) \|Y_0^T X_0\| \epsilon}$$

Or on a

$$\begin{aligned} \|\mathbb{P}_0 - \hat{\mathbb{P}}_0^{(n)}\| &\leq \|X_0 - \hat{X}_0^{(n)}\| \|(Y_0^T X_0)^{-1} Y_0^T\| + \|\hat{X}_0^{(n)} (Y_0^T X_0)^{-1}\| \|Y_0 - \hat{Y}_0^{(n)}\| \\ &\quad + \|\hat{X}_0^{(n)}\| \left\| (Y_0^T X_0)^{-1} - \left( (\hat{Y}_0^{(n)})^T \hat{X}_0^{(n)} \right)^{-1} \right\| \|\hat{Y}_0^{(n)}\| \\ &\leq \|(Y_0^T X_0)^{-1}\| (\|X_0 - \hat{X}_0^{(n)}\| + \|Y_0 - \hat{Y}_0^{(n)}\|) \\ &\quad + \|(Y_0^T X_0)^{-1}\| \|Y_0^T X_0 - (\hat{Y}_0^{(n)})^T \hat{X}_0^{(n)}\| \left\| \left( (\hat{Y}_0^{(n)})^T \hat{X}_0^{(n)} \right)^{-1} \right\|. \end{aligned}$$

La démonstration se termine en utilisant les différentes majorations démontrées ci-dessus. On obtient alors

$$\begin{aligned} \|\mathbb{P}_0 - \hat{\mathbb{P}}_0^{(n)}\| &\leq \|(Y_0^T X_0)^{-1}\| \left( 4 \|Y_0^T X_0\| \epsilon + 5 \frac{\|Y_0^T X_0\|^2 \|(Y_0^T X_0)^{-1}\|}{1 - 5 \kappa_2 (Y_0^T X_0) \|Y_0^T X_0\| \epsilon} \epsilon \right) \\ &\leq \kappa_2 (Y_0^T X_0) \left( 4 + 5 \frac{\kappa_2 (Y_0^T X_0)}{1 - 5 \kappa_2 (Y_0^T X_0) \|Y_0^T X_0\| \epsilon} \right) \epsilon \\ &\leq \kappa_2 (Y_0^T X_0) \left( 4 + 5 \frac{\kappa_2 (Y_0^T X_0)}{1 - 5 \kappa_2 (Y_0^T X_0) \epsilon} \right) \epsilon. \end{aligned}$$

□

Le théorème 15 montre qu'on peut construire à partir de  $\tilde{X}_0^{(n)}$  un projecteur qui est une approximation de  $\mathbb{P}_0$ . Il est clair qu'en appliquant la même méthode, on peut déterminer un projecteur  $\hat{\mathbb{P}}_\infty^{(n)}$  qui approche  $\mathbb{P}_\infty$ . Or les projecteurs  $\mathbb{P}_0$ ,  $\mathbb{P}_1$  et  $\mathbb{P}_\infty$  vérifient

$$\begin{aligned}\mathbb{P}_0 + \mathbb{P}_1 + \mathbb{P}_\infty &= I_{2N} \\ \mathbb{P}_0 \mathbb{P}_1 &= 0, \\ \mathbb{P}_0 \mathbb{P}_\infty &= 0, \\ \mathbb{P}_1 \mathbb{P}_0 &= 0, \\ \mathbb{P}_1 \mathbb{P}_\infty &= 0, \\ \mathbb{P}_\infty \mathbb{P}_0 &= 0, \\ \mathbb{P}_\infty \mathbb{P}_1 &= 0.\end{aligned}$$

Ces relations offrent la possibilité de vérifier la qualité numérique de  $\hat{\mathbb{P}}_0^{(n)}$ ,  $\tilde{\mathbb{P}}_1^{(n)}$  et  $\hat{\mathbb{P}}_\infty^{(n)}$ . En calculant les erreurs suivantes :

$$\begin{aligned}E &= \left\| I_{2N} - (\hat{\mathbb{P}}_0^{(n)} + \tilde{\mathbb{P}}_1^{(n)} + \hat{\mathbb{P}}_\infty^{(n)}) \right\|, \\ E_{0,1} &= \left\| \hat{\mathbb{P}}_0^{(n)} \tilde{\mathbb{P}}_1^{(n)} \right\|, \\ E_{0,\infty} &= \left\| \hat{\mathbb{P}}_0^{(n)} \hat{\mathbb{P}}_\infty^{(n)} \right\|, \\ E_{1,0} &= \left\| \tilde{\mathbb{P}}_1^{(n)} \hat{\mathbb{P}}_0^{(n)} \right\|, \\ E_{1,\infty} &= \left\| \tilde{\mathbb{P}}_1^{(n)} \hat{\mathbb{P}}_\infty^{(n)} \right\|, \\ E_{\infty,0} &= \left\| \hat{\mathbb{P}}_\infty^{(n)} \hat{\mathbb{P}}_0^{(n)} \right\|, \\ E_{\infty,1} &= \left\| \hat{\mathbb{P}}_\infty^{(n)} \tilde{\mathbb{P}}_1^{(n)} \right\|.\end{aligned}$$

On dispose ainsi de plusieurs indicateurs permettant de valider, en partie, la qualité des différents projecteurs calculés par l'algorithme de trichotomie dans le sens où si un de ces indicateurs n'est pas "suffisamment" petit alors il existe un doute légitime quant aux quantités calculées.

Considérer toutes les quantités  $E$ ,  $E_{0,1}$ ,  $E_{1,0}$ ,  $E_{0,\infty}$ ,  $E_{\infty,0}$ ,  $E_{1,\infty}$  et  $E_{\infty,1}$  peut sembler, à première vue, redondant. En effet, par exemple si  $E$ ,  $E_{0,1}$ ,  $E_{1,0}$ ,  $E_{0,\infty}$ ,  $E_{\infty,0}$  et  $E_{1,\infty}$  sont nulles alors  $E_{\infty,1}$  l'est également car, dans

cette situation, on a

$$\begin{aligned}
0 &= I_{2N} - \widehat{\mathbb{P}}_0^{(n)} + \widetilde{\mathbb{P}}_1^{(n)} + \widehat{\mathbb{P}}_\infty^{(n)} \\
&= \left( I_{2N} - \widehat{\mathbb{P}}_0^{(n)} + \widetilde{\mathbb{P}}_1^{(n)} + \widehat{\mathbb{P}}_\infty^{(n)} \right) \left( \widehat{\mathbb{P}}_0^{(n)} + \widetilde{\mathbb{P}}_1^{(n)} + \widehat{\mathbb{P}}_\infty^{(n)} \right) \\
&= \widehat{\mathbb{P}}_0^{(n)} + \widetilde{\mathbb{P}}_1^{(n)} + \widehat{\mathbb{P}}_\infty^{(n)} - \widehat{\mathbb{P}}_0^{(n)}\widehat{\mathbb{P}}_0^{(n)} - \widetilde{\mathbb{P}}_1^{(n)}\widehat{\mathbb{P}}_0^{(n)} - \widehat{\mathbb{P}}_\infty^{(n)}\widehat{\mathbb{P}}_0^{(n)} - \widehat{\mathbb{P}}_0^{(n)}\widetilde{\mathbb{P}}_1^{(n)} \\
&\quad - \widetilde{\mathbb{P}}_1^{(n)}\widetilde{\mathbb{P}}_1^{(n)} - \widehat{\mathbb{P}}_\infty^{(n)}\widetilde{\mathbb{P}}_1^{(n)} - \widehat{\mathbb{P}}_0^{(n)}\widehat{\mathbb{P}}_\infty^{(n)} - \widetilde{\mathbb{P}}_1^{(n)}\widehat{\mathbb{P}}_\infty^{(n)} - \widehat{\mathbb{P}}_\infty^{(n)}\widehat{\mathbb{P}}_\infty^{(n)} \\
&= \widehat{\mathbb{P}}_0^{(n)} + \widetilde{\mathbb{P}}_1^{(n)} + \widehat{\mathbb{P}}_\infty^{(n)} - \widehat{\mathbb{P}}_0^{(n)} - \widetilde{\mathbb{P}}_1^{(n)}\widehat{\mathbb{P}}_0^{(n)} - \widehat{\mathbb{P}}_\infty^{(n)}\widehat{\mathbb{P}}_0^{(n)} - \widehat{\mathbb{P}}_0^{(n)}\widetilde{\mathbb{P}}_1^{(n)} \\
&\quad - \widetilde{\mathbb{P}}_1^{(n)} - \widehat{\mathbb{P}}_\infty^{(n)}\widetilde{\mathbb{P}}_1^{(n)} - \widehat{\mathbb{P}}_0^{(n)}\widehat{\mathbb{P}}_\infty^{(n)} - \widetilde{\mathbb{P}}_1^{(n)}\widehat{\mathbb{P}}_\infty^{(n)} - \widehat{\mathbb{P}}_\infty^{(n)} \\
&= -\widehat{\mathbb{P}}_\infty^{(n)}\widetilde{\mathbb{P}}_1^{(n)}.
\end{aligned}$$

Cependant dans le cas général, cela n'est pas aussi simple, en particulier, en ce qui concerne les couples  $E_{0,1}$  et  $E_{1,0}$ ,  $E_{0,\infty}$  et  $E_{\infty,0}$ ,  $E_{1,\infty}$  et  $E_{\infty,1}$ . En effet, l'étape de  $J$ -orthogonalisation utilisée pour éliminer (du moins réduire) les composantes associées à  $\mathcal{X}_1$  dans  $\mathcal{X}_0^{(n)}$  et  $\mathcal{X}_\infty^{(n)}$  permet d'assurer que  $E_{1,0}$  et  $E_{1,\infty}$  seront nécessairement petites, cela même si les suites  $(X_0^{(n)})_{n \geq 0}$ ,  $(X_1^{(n)})_{n \geq 0}$  et  $(X_\infty^{(n)})_{n \geq 0}$  sont loin d'avoir convergé. Par exemple, sachant que

$$\widetilde{X}_0^{(n)} = \left( I - \widetilde{\mathbb{P}}_1^{(n)} \right) X_0^{(n)}$$

on vérifie que  $\widetilde{\mathbb{P}}_1^{(n)}\widetilde{X}_0^{(n)} = 0$  et donc, par construction,  $\widetilde{\mathbb{P}}_1^{(n)}\widehat{\mathbb{P}}_0^{(n)} = 0$ . Cependant si  $E_{1,0}$  et  $E_{1,\infty}$  sont théoriquement nulles, il peut être tout de même intéressant de les calculer car en pratique, elles ne seront pas nulles à cause des erreurs d'arrondies. Ces quantités dépendent de l'erreur machine (ou  $\epsilon$ -machine) et du conditionnement des sous-espaces invariants  $\mathcal{X}_0$ ,  $\mathcal{X}_1$  et  $\mathcal{X}_\infty$ , en supposant que  $\widehat{\mathcal{X}}_0^{(n)}$ ,  $\widetilde{\mathcal{X}}_1^{(n)}$  et  $\widehat{\mathcal{X}}_\infty^{(n)}$  en sont suffisamment proches. Elles offrent donc la possibilité d'estimer le conditionnement spectral de  $W$ , autrement dit la difficulté du problème et ainsi de relativiser les autres estimateurs.

**Remarque 14** *Il est important de noter que si ces estimateurs peuvent permettre d'estimer la qualité des projecteurs calculés par l'algorithme de trichotomie, ils ne possèdent pas la fiabilité du critère de dichotomie représenté par la norme de la matrice  $H$  du Chapitre 2.*

Une autre possibilité consiste à définir  $\widehat{\mathbb{P}}_\infty^{(n)}$  par

$$\widehat{\mathbb{P}}_\infty^{(n)} = I_{2N} - \widehat{\mathbb{P}}_0^{(n)} - \widetilde{\mathbb{P}}_1^{(n)}.$$

Cela permet de s'assurer que l'estimateur  $E$  est nul. Cependant ainsi défini,  $\widehat{\mathbb{P}}_\infty^{(n)}$  n'est, a priori, pas un projecteur. Or l'algorithme de trichotomie a pour

but de calculer les projecteurs et donc cette définition ne présente que peu d'intérêt.

### 3.6.3 Algorithme de trichotomie

L'algorithme de trichotomie est donné dans son entier.

#### Algorithme 4

1. Résoudre

$$\begin{pmatrix} JW & -W^T J & 0 \\ 0 & JW & -W^T J \\ \alpha I & \beta I & \alpha I \end{pmatrix} \begin{pmatrix} K_0^{(1)} \\ K_1^{(1)} \\ K_\infty^{(1)} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ I \end{pmatrix}.$$

2. Poser

- $X_0^{(1)} = K_0^{(1)},$
- $X_1^{(1)} = K_1^{(1)},$
- $X_\infty^{(1)} = K_\infty^{(1)}.$

3. Pour  $j = 2$  à  $n$ ,

- Résoudre

$$\begin{pmatrix} \left(X_1^{(j-1)}\right)^T J & -\left(X_\infty^{(j-1)}\right)^T J & 0 \\ 0 & \left(X_0^{(j-1)}\right)^T J & -\left(X_1^{(j-1)}\right)^T J \\ \alpha X_0^{(j-1)} & \beta X_1^{(j-1)} & \alpha X_\infty^{(j-1)} \end{pmatrix} \begin{pmatrix} K_0^{(j)} \\ K_1^{(j)} \\ K_\infty^{(j)} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ I \end{pmatrix}.$$

- Poser

- .  $X_0^{(j)} = X_0^{(j-1)} K_0^{(j)},$
- .  $X_1^{(j)} = X_1^{(j-1)} K_1^{(j)},$
- .  $X_\infty^{(j)} = X_\infty^{(j-1)} K_\infty^{(j)}.$

4. Calculer la décomposition en valeurs singulières de  $X_1^{(n)}$

$$X_1^{(n)} = U \Sigma V^T.$$

- A la vue des valeurs singulières, déduire une estimation  $N_1$ .
- Poser  $\tilde{X}_1^{(n)} = U_1$  où  $U_1$  est la matrice formée des  $2N_1$  premières colonnes de  $U$ .

- Calculer  $\tilde{\mathbb{P}}_1^{(n)}$  par

$$\tilde{\mathbb{P}}_1^{(n)} = \tilde{X}_1^{(n)} \left( \left( \tilde{X}_1^{(n)} \right)^T J \tilde{X}_1^{(n)} \right)^{-1} \left( \tilde{X}_1^{(n)} \right)^T J.$$

5. Calculer  $\tilde{\mathbb{P}}_0^{(n)}$  et  $\tilde{\mathbb{P}}_\infty^{(n)}$  par

$$\tilde{\mathbb{P}}_0^{(n)} = \alpha \left( I - \tilde{\mathbb{P}}_1^{(n)} \right) X_0^{(n)},$$

$$\tilde{\mathbb{P}}_\infty^{(n)} = \alpha \left( I - \tilde{\mathbb{P}}_1^{(n)} \right) X_\infty^{(n)}.$$

Le tableau suivant donne le coût global de l'algorithme 4.

<hr/>	
étapes 1, 2, 3	
voir le tableau 3.3	$(38n - 6) (2N)^3$
étape 4	
Décomposition SVD (voir [13])	$8/3(2N)^3$
Calcul de $\tilde{\mathbb{P}}_1^{(n)}$	$(2/3(2N_1)^2 + 5(2N_1)(2N) + 2(2N)^2) (2N_1)$
étape 5	
Calcul de $\tilde{\mathbb{P}}_0^{(n)}$	$2(2N)^3$
Calcul de $\tilde{\mathbb{P}}_\infty^{(n)}$	$2(2N)^3$
<hr/>	
Total	$2/3(2N_1)^3 + 5(2N_1)^2(2N) + 2(2N_1)(2N)^2 + (38n + 2/3) (2N)^3$

TAB. 3.4 – Coût des opérations de l'algorithme 4

**Remarque 15** Rappelons que le coût total de  $n$  itérations de l'algorithme 2 appliqué à  $W$  est

$$\left( \frac{26}{3} + \frac{61}{3}n \right) (2N)^3.$$

Rappelons également qu'il faut le portrait spectral pour déterminer  $r_0$  et deux applications de l'algorithme 2 pour obtenir  $P_{r_0}$  et  $P_{1/r_0}$  des approximations de  $\mathbb{P}_0$  et  $I - \mathbb{P}_\infty$ . Ceci rend l'approche "trichotomie spectrale" de ce chapitre généralement beaucoup plus économique que l'approche "dichotomie spectrale" sans toutefois posséder la fiabilité de cette dernière. En effet, nous n'avons pas réussi à trouver l'équivalent du critère de dichotomie pour l'algorithme 4.

### 3.7 Une autre approche de détermination de la stabilité forte

L'application de l'algorithme 4 à la matrice  $J$ -symplectique  $W$  nous permet de déterminer les projecteurs  $\mathbb{P}_0$ ,  $\mathbb{P}_1$  et  $\mathbb{P}_\infty$ . Lorsque  $\mathbb{P}_0 = \mathbb{P}_\infty = 0$  et donc  $\mathbb{P}_1 = I$ , on peut soit reprendre l'algorithme 3 du chapitre 2, soit utiliser l'approche développée dans [6] pour analyser la stabilité et la stabilité forte de  $W$ .

Cette approche n'utilise pas de dichotomie, elle ne fait intervenir que des calculs spectraux sur des matrices symétriques pour lesquelles le calcul se fait avec une précision garantie (voir [21]). Nous en résumons les étapes principales : elle commence par calculer la suite matricielle  $(S^{(n)})_{n \geq 0}$  définie par :

$$S^{(n)} = \frac{1}{2^n} \sum_{k=1}^{2^n} (W^T)^{k-1} W^{k-1}.$$

Remarquons que  $S^{(n)}$  est symétrique définie positive :  $S^{(n)} = (S^{(n)})^T > 0$ . Son calcul s'effectue de manière itérative comme suit :

$$\begin{aligned} S^{(0)} &= I, \quad B^{(0)} = W, \quad B^{(n)} = \left(B^{(n-1)}\right)^2, \\ S^{(n)} &= \frac{1}{2} \left( S^{(n-1)} + \left(B^{(n-1)}\right)^T S^{(n-1)} B^{(n-1)} \right). \end{aligned}$$

Si  $W$  est stable alors cette suite converge vers une matrice  $S^{(\infty)} = (S^{(\infty)})^T > 0$  satisfaisant

$$W^T S^{(\infty)} W = S^{(\infty)}.$$

Pour analyser la stabilité forte, on considère le faisceau symétrique  $S_0 - \lambda S^{(\infty)}$  avec  $S_0 = 1/2 J(W - W^{-1}) = 1/2(JW + (JW)^T)$ . En désignant par  $\mathbb{P}_-$  et  $\mathbb{P}_+$  les projecteurs spectraux sur les s-e-i associés aux valeurs propres négatives et positives de  $S_0 - \lambda S^{(\infty)}$ , alors la stabilité forte est caractérisée par :

$$\begin{aligned} \mathbb{P}_- + \mathbb{P}_+ &= I_{2N}, \quad \mathbb{P}_\pm W = W \mathbb{P}_\pm \\ \mathbb{P}_-^T S_0 \mathbb{P}_- &\leq 0, \quad \mathbb{P}_+^T S_0 \mathbb{P}_+ \geq 0 \\ S &= S_0(\mathbb{P}_+ - \mathbb{P}_-) = S^T > 0. \end{aligned}$$

D'autre part, soit  $V = S^{\frac{1}{2}} W S^{-\frac{1}{2}}$ . Alors  $V$  est orthogonale et vérifie  $W^T S W = S$ . En considérant la matrice symétrique  $C = (V + V^T)/2$  dont les valeurs

propres sont de la forme  $\cos \omega_j$  où  $e^{\pm i\omega_j}$  sont des valeurs propres de  $W$ , alors la forme quadratique  $(S_0x, x)$  doit être non nulle et de signe constant pour tout vecteur propre  $x$  de  $C$ .

Si toutes ces conditions sont remplies, alors  $\mathbb{P}_r = \mathbb{P}_+$  et  $\mathbb{P}_v = \mathbb{P}_-$ . Voir [6] pour les détails.

Nous résumons cette approche dans l'algorithme ci-dessous. L'étape préalable à l'analyse de la stabilité consiste à s'assurer d'abord que  $\mathbb{P}_0 = \mathbb{P}_\infty = 0$ ,  $\mathbb{P}_1 = I$ . Cette étape est effectuée par l'algorithme 4.

### Algorithme 5

1. Déterminer les projecteurs  $\mathbb{P}_0$ ,  $\mathbb{P}_1$  et  $\mathbb{P}_\infty$  par l'algorithme 4<sup>1</sup>.
2. Si  $\mathbb{P}_0 \neq 0$  ( ou  $\mathbb{P}_\infty \neq 0$  ), il n'y a pas stabilité (forte). Sinon calculer la suite  $S^{(n)}$ . Si  $S^{(n)}$  ne converge pas, il n'y a pas stabilité (forte). Sinon calculer la matrice

$$S_0 = (1/2)((JW) + (JW)^T)$$

et le projecteur spectral  $\mathbb{P}_-$  sur le s-e-i associé aux valeurs propres négatives du faisceau  $S_0 - \lambda S^{(\infty)}$ , où  $S^{(\infty)}$  est "la limite" de la suite  $S^{(n)}$ , puis calculer

$$S = S_0(I - 2\mathbb{P}_-)$$

3. Si  $S$  est symétrique définie positive, calculer

$$V = S^{1/2}WS^{-1/2} \quad \text{et} \quad C = (1/2)(V + V^T).$$

Si la forme quadratique  $(S_0x, x)$  est non nulle et de signe constant pour tout vecteur propre  $x$  de  $C$ , alors il y a stabilité forte. Dans ce cas

$$\mathbb{P}_r = I - \mathbb{P}_- \quad \text{et} \quad \mathbb{P}_v = \mathbb{P}_-$$

## 3.8 Illustrations numériques

### Exemple 1

On applique la méthode étudiée dans ce chapitre à la matrice symplectique

---

<sup>1</sup>Pour ne pas alourdir les notations,  $\mathbb{P}_0$ ,  $\mathbb{P}_1$  et  $\mathbb{P}_\infty$  désignent ici des approximations des projecteurs exacts calculées par l'algorithme 4



de l'exemple 4. Considérons la matrice

$$B = \begin{pmatrix} 0.3473 & -0.0452 & 0.3832 & -0.8490 & -0.0973 & -0.0181 \\ 0.3933 & -0.4122 & -0.4396 & -0.0900 & 0.6753 & -0.1343 \\ 0.3460 & 0.7347 & 0.1263 & 0.1036 & 0.4173 & 0.3738 \\ 0.4356 & -0.3794 & 0.6508 & 0.4924 & 0.0010 & -0.0150 \\ 0.4236 & -0.1760 & -0.4036 & 0.0431 & -0.4874 & 0.6223 \\ 0.4856 & 0.3367 & -0.2398 & 0.1270 & -0.3505 & -0.6741 \end{pmatrix}$$

de valeurs propres  $0.7456 \pm 0.6664i$ ,  $-0.2804 \pm 0.9599i$  et  $-0.7690 \pm 0.6392i$  toutes de module 1. Alors la matrice  $W$  suivante est  $J$ -symplectique

$$W = \begin{pmatrix} B & \\ & B^{-T} \end{pmatrix} \quad \text{où} \quad J = \begin{pmatrix} & -I \\ I & \end{pmatrix}$$

et a toutes ses valeurs propres de module 1 et différentes de  $\pm 1$ . L'application de l'algorithme 4 à cette matrice donne  $\mathbb{P}_0 = \mathbb{P}_\infty = 0$  et  $\mathbb{P}_1 = I$ . Le calcul de la suite  $S^{(n)}$  nous donne  $S^{(n)} = I$ ,  $\forall n$  ( $W$  étant orthogonale,  $(W^{k-1})^T W^{k-1} = I$ ,  $\forall k \geq 1$ ). Après  $n_0 = 30$  itérations, on a une approximation  $S^{(n_0)}$  de  $S^{(\infty)}$  satisfaisant à  $\|WS^{(n_0)}W - S^{(n_0)}\| = 4.5370e - 015$ . On vérifie que les projecteurs spectraux  $\mathbb{P}_-$  et  $\mathbb{P}_+$  sur les sous-espaces invariants du faisceau  $S_0 - \lambda S^{(\infty)}$  associés respectivement aux valeurs propres négatives et positives sont tels que

$$\|\mathbb{P}_-^2 - \mathbb{P}_-\| = 3.4980e - 016 \quad \text{et} \quad \|\mathbb{P}_+^2 - \mathbb{P}_+\| = 3.4605e - 016$$

avec  $\mathbb{P}_+ = I - \mathbb{P}_-$ . Ces deux projecteurs sont de trace 6.

Posons  $X = \mathbb{P}_+ - \mathbb{P}_- = I - 2\mathbb{P}_-$  et  $S = S_0 X$ ; alors  $S$  est symétrique ( $\|S - S^T\| = 1.5708e - 015$ ), définie positive (ses valeurs propres sont 0.9599, 0.6664 et 0.6392). En considérant la matrice orthogonale  $V = S^{1/2} W S^{-1/2}$ , puis la matrice symétrique  $C = (V + V^T)/2$ , la forme quadratique  $(S_0 x, x)$  n'est pas de signe constant pour tout vecteur propre  $x$  de  $C$ . En effet, il existe deux vecteurs propres  $x_1$  et  $x_2$  associés à la valeur propre double  $\lambda = -0.2804$  de  $C$  tels que  $(S_0 x_1, x_2) = 0.4683$  et  $(S_0 x_2, x_2) = -0.4515$ . Donc  $W$  n'est pas pas fortement stable. Autrement dit on a la même conclusion qu'au chapitre 2.

## Exemple 2

On reprend trois cas de l'exemple 5 du chapitre 2.

- a)  $t = 0.39050$ , l'application de l'algorithme 4 à  $W$  et  $J$ , donne  $\mathbb{P}_0 \neq 0$ ,  $\mathbb{P}_\infty \neq 0$  et  $\mathbb{P}_1 = 0$  avec

$$\|\mathbb{P}_0^2 - \mathbb{P}_0\| = 1.0231e - 013 \quad \text{et} \quad \|\mathbb{P}_\infty^2 - \mathbb{P}_\infty\| = 5.8870e - 014.$$

Les erreurs commises dans le calcul de ces projecteurs sont :

$$E_{0,\infty} = 2.1199e - 014 \quad \text{et} \quad E_{\infty,0} = 1.9928e - 014.$$

Ces erreurs montrent que les projecteurs sont bien calculés. Les deux projecteurs non nuls sont de trace 2. Aucune valeur propre n'est sur le cercle unité.  $W$  n'est pas (fortement) stable.

- b)  $t = 0.3896$  l'application de l'algorithme 4 à  $W$  et  $J$  donne  $\mathbb{P}_0 = \mathbb{P}_\infty = 0$  et  $\mathbb{P}_1 = I$ .

Le calcul de la suite  $S^{(n)}$  s'arrête lorsque  $\frac{\|S^{(n_0+1)} - S^{(n_0)}\|}{\|S^{(n_0)}\|} < 10^{-5}$ . Après  $n_0 = 24$  itérations, on a une approximation  $S^{(n_0)}$  de  $S^{(\infty)}$  satisfaisant à  $\|W^T S^{(n_0)} W - S^{(n_0)}\| = 2.5965e - 007$ .

On vérifie que les projecteurs spectraux  $\mathbb{P}_-$  et  $\mathbb{P}_+$  sur les sous-espaces invariants du faisceau  $S_0 - \lambda S^{(\infty)}$  associés respectivement aux valeurs propres négatives et positives sont tels que :  $\mathbb{P}_+ = I - \mathbb{P}_-$  avec

$$\|\mathbb{P}_-^2 - \mathbb{P}_-\| = 1.2654e - 013 \quad \text{et} \quad \|\mathbb{P}_+^2 - \mathbb{P}_+\| = 1.2654e - 013.$$

Ces deux projecteurs sont de trace 2.

Posons  $X = \mathbb{P}_+ - \mathbb{P}_- = I - 2\mathbb{P}_-$  et  $S = S_0 X$ . Alors la matrice  $S$  est symétrique ( $\|S - S^T\| = 5.3577e - 012$ ) et définie positive (ses valeurs propres sont 26.4893, 17.2845, 0.0317, 0.0205).

En considérant la matrice orthogonale  $V = S^{1/2} W S^{-1/2}$  et la matrice symétrique  $C = (V + V^T)/2$ , la forme quadratique  $(S_0 x, x)$  est de signe constant pour toute vecteur propre  $x$  associée aux valeurs propres de  $C$ . Donc  $W$  est fortement stable et

$$\mathbb{P}_r = \mathbb{P}_+ \quad \text{et} \quad \mathbb{P}_v = \mathbb{P}_-.$$

On a les mêmes résultats qu'au chapitre 2.

- c)  $t = 0.39$ , l'application de l'algorithme 4 à  $W$  et  $J$  donne comme on s'y attend  $\mathbb{P}_0 = \mathbb{P}_\infty = 0$  et  $\mathbb{P}_1 = I$ .

Le calcul de la suite  $S^{(n)}$  s'arrête lorsque  $\frac{\|S^{(n_0+1)} - S^{(n_0)}\|}{\|S^{(n_0)}\|} < 10^{-5}$ . Après  $n_0 = 29$  itérations, on a une approximation  $S^{(n_0)}$  de  $S^{(\infty)}$  satisfaisant à  $\|W^T S^{(n_0)} W - S^{(n_0)}\| = 1.9397e - 006$ .

On vérifie que les projecteurs spectraux  $\mathbb{P}_-$  et  $\mathbb{P}_+$  sur les sous-espaces invariants du faisceau  $S_0 - \lambda S^{(\infty)}$  associés respectivement aux valeurs propres négatives et positives sont tels que :  $\mathbb{P}_+ = I - \mathbb{P}_-$  avec

$$\|\mathbb{P}_-^2 - \mathbb{P}_-\| = 1.2814e - 011 \quad \text{et} \quad \|\mathbb{P}_+^2 - \mathbb{P}_+\| = 1.2814e - 011.$$

*Ces deux projecteurs sont de trace 2.*

*Posons  $X = \mathbb{P}_+ - \mathbb{P}_-$  et  $S = S_0 X$  ; alors  $S$  est symétrique ( $\|S - S^T\| = 1.5757e - 009$ ) et définie positive (ses valeurs propres sont 129.4300, 84.0936, 0.0065). En considérant la matrice orthogonale  $V = S^{1/2} W S^{-1/2}$  et la matrice symétrique  $C = (V + V^T)/2$ , la forme quadratique  $(S_0 x, x)$  est de signe constant pour tout vecteur propre  $x$  de  $C$ . Donc  $W$  est fortement stable et*

$$\mathbb{P}_r = \mathbb{P}_+ \quad \text{et} \quad \mathbb{P}_v = \mathbb{P}_-.$$

*On a les mêmes conclusions qu'au chapitre 2.*

### 3.9 Conclusion

Dans ce chapitre, nous avons proposé une approche permettant le calcul des projecteurs spectraux  $\mathbb{P}_0$ ,  $\mathbb{P}_1$  et  $\mathbb{P}_\infty$  sur les s-e-i associés respectivement aux valeurs propres intérieures, sur et extérieures au cercle unité d'une matrice symplectique. Cette approche, comme celle développée au chapitre 2, est utilisée dans ce travail comme une étape préalable à l'analyse de la stabilité forte d'une matrice symplectique (cela permet de s'assurer d'abord que  $\mathbb{P}_1 = I$ ,  $\mathbb{P}_0 = \mathbb{P}_\infty = 0$ ). Bien entendu, elle peut être utilisée dans d'autres contextes, par exemple, dans le domaine du contrôle optimal on a  $\mathbb{P}_1 = 0$ ,  $\mathbb{P}_0 \neq 0$ ,  $\mathbb{P}_\infty \neq 0$  (voir par exemple [12]).

Cette approche présente un avantage certain par rapport à celle basée sur la dichotomie spectrale du chapitre 2 : elle est moins coûteuse en nombre d'opérations arithmétiques. Cependant, l'approche basée sur la dichotomie spectrale possède des qualités théorique et numérique assez robustes. Un travail futur consistera à faire une analyse comparative de ces deux approches, et particulier : i) comparer les qualités des projecteurs  $\mathbb{P}_0$ ,  $\mathbb{P}_1$  et  $\mathbb{P}_\infty$  obtenus par les deux approches, ii) analyser la stabilité numérique de l'algorithme 4. D'autre part, il serait intéressant d'utiliser les propriétés de symplecticité de manière intensive dans l'algorithme 4, en particulier dans l'étape 4 de cet algorithme. Enfin, l'étude de la stabilité forte par les algorithmes 3 et 5 mérite une comparaison approfondie.

# Bibliographie

- [1] Z. Bai, J. Demmel, M. Gu, Inverse free parallel spectral divide and conquer algorithms for nonsymmetric eigenproblems, *Numerische Mathematik* 76 (1997) 279-308.
- [2] C. Bekas, E. Gallopoulos, Cobra : parallel path following for computing the matrix pseudospectrum, *Parallel Comput.*, 27 (2001) 1879-1896.
- [3] C. Brezinski, *Computational Aspects of Linear Control*, Kluwer Academic Publishers, 2002.
- [4] S. K. Godunov, Problem of the dichotomy of the spectrum of a matrix, *Siberian Math. J.* 27 (1986) 649-660.
- [5] S. K. Godunov, Stability of iterations of symplectic transformations, *Siberian Math. J.* 30 (1989) 54-63.
- [6] S. K. Godunov, Verification of boundedness for the powers of symplectic matrices with the help of averaging, *Siberian Math. J.*, 33 (1992) 939-949.
- [7] S.K. Godunov, M. Sadkane. Elliptic dichotomy of a matrix spectrum, *Linear Algebra Appl.* 248 (1996) 205-232.
- [8] S. K. Godunov, *Modern Aspects of Linear Algebra*, Amer. Math. Soc., Vol. 175, 1998.
- [9] S.K. Godunov, M. Sadkane. Computation of pseudospectra via spectral projectors, *Linear Algebra Appl.* 279 (1998) 163-175.
- [10] S. K. Godunov, M. Sadkane, Numerical determination of a canonical form of a symplectic matrix, *Siberian Math. J.* 42 (2001) 629-647.
- [11] S. K. Godunov, M. Sadkane, Some new algorithms for the spectral dichotomy methods, *Linear Algebra Appl.* 358 (2003) 173-194.
- [12] S. K. Godunov, M. Sadkane, Analysis of symplectic matrices by their spectral portraits. Applications to optimal control and to the theory of parametric resonance, Manuscript 2004.

- [13] G. H. Golub, F. Van Loan, Matrix computations, 2nd ed., The Johns Hopkins University Press, Baltimore, MD, 1989.
- [14] B. Hassibi, A. H. Sayed, T. Kailath, Indefinite-Quadratic Estimation and Control, SIAM, Philadelphia, PA, 1999.
- [15] T. Kato, Perturbation theory for linear operators, Springer-Verlag, New York, 1976.
- [16] P. Lancaster, L. Rodman, Algebraic Riccati Equations, Clarendon Press, 1995.
- [17] P.-F. Lavallée, M. Sadkane, Computation of pseudospectra by spectral dichotomy methods in a parallel environment. Numerical Algorithms 33 (2003) 343-355.
- [18] A. N. Malyshev, Guaranteed Accuracy in Spectral Problems of Linear Algebra, Siberian Adv. Math. J. I,II 2 (1992) 144-197.
- [19] A.N. Malyshev, Parallel algorithm for solving some spectral problems of linear algebra, Linear Algebra Appl. 188-189 (1993) 489-520.
- [20] A. N. Malyshev, M. Sadkane, On parabolic and elliptic spectral dichotomy, SIAM J. Matrix Anal. Applic. 18 (1997) 267-278.
- [21] B.N. Parlett, The symmetric eigenvalue problem, Classics in Applied Mathematics, SIAM, Philadelphia, 1988.
- [22] M. Robbé, M. Sadkane, Riccati-based preconditioner for computing invariant subspaces of large matrices, Numerische Mathematik 92 (2002) 129-159.
- [23] A. Salam, On theoretical and numerical aspects of symplectic Gram-Schmidt-like algorithms, Numerical Algorithms 39 (2005) 437-462.
- [24] M. Sadkane, Estimates from the discrete-time Lyapunov equation, Appl. Math. Lett. 16 (2003) 313-316.
- [25] M. Sadkane, Norm estimates of the Fourier series coefficients of the matrix resolvent, Appl. Math. Lett. 18 (2005) 149-153.
- [26] G.W. Stewart, J.-G. Sun, Matrix Perturbation Theory, Academic Press, San Diego, CA, 1990.
- [27] L. N. Trefethen, M. Embree, Spectra and pseudospectra. The behavior of non-normal matrices and operators, Princeton University Press Princeton, 2005.
- [28] H. Xu, An SVD-Like Matrix Decomposition and Its Applications, Linear Algebra Appl., 368 (2003) 1-24.
- [29] V.A. Yakubovich, V.M. Starzhinskii, Linear differential equations with periodic coefficients, Vol. 1 & 2., Wiley, New York, 1975.

## Sur quelques algorithmes d'analyse de la stabilité forte de matrices symplectiques.

**Résumé :** Le but de cette thèse est de proposer quelques algorithmes permettant d'analyser la stabilité forte d'une matrice symplectiques.

Le premier chapitre rassemble quelques propriétés spectrales des matrices symplectiques et leur lien avec la stabilité forte. En particulier, ce chapitre insiste sur les propriétés les plus importantes du point de vue de la stabilité numérique qui sont utilisées tout au long de la thèse.

Dans le deuxième chapitre, on adapte les méthodes de dichotomie spectrale à des matrices symplectiques. Cette adaptation permet de trouver les projecteurs spectraux sur les sous-espaces invariants associés aux valeurs propres dans, sur et en dehors du cercle unité. Puis, on propose un algorithme basé encore sur la dichotomie spectrale afin d'analyser la stabilité forte.

Dans le troisième chapitre, on propose un algorithme itératif de trichotomie spectrale permettant de trouver les projecteurs spectraux mentionnés ci-dessus. L'analyse théorique clarifie la convergence et donne la vitesse de convergence de cet algorithme. Ce chapitre se termine par une comparaison, en termes de précision et de coût de calcul, entre cette approche et celle du chapitre 2.

**Mots clés :** matrice symplectique, stabilité forte, dichotomie spectrale

## On some algorithms for the strong stability analysis of symplectic matrices

**Abstract :** The first chapter of this thesis presents some spectral properties of symplectic matrices and their link to strong stability. In particular, emphasis is placed on the most numerically useful properties which are used throughout the thesis.

In the second chapter we propose an adaptation of spectral dichotomy algorithms to symplectic matrices. This adaptation allows us to find, in a stable way, the spectral projections onto the invariant subspaces associated with the eigenvalues inside, on and outside the unit circle. With these information, we propose an algorithm, based again on the spectral dichotomy techniques, to analyze the strong stability.

In the third chapter, we propose another approach referred to as spectral trichotomy, for computing the abovementioned spectral projections. We analyse the convergence behavior of this algorithm and compare it with the approach proposed in chapter 2.

**Key words :** symplectic matrix, strong stability, spectral dichotomy